

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 7.85318 Seconds
(without alignments)
3506.895 Million cell updates/sec

Title: US-09-856-221-4
Perfect score: 605
Sequence: 1 gtgaagcggcagtgatgcaaa.....atttggaagaagatgaacgg 332

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-O/cqn2_1/USPTO.spool/US09856221/runat_15012003_153924_12445/app_query_fasta_1.1948
-DB=SwissProt40 -QFW=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221 -ECN_1_1_38 -runat_15012003_153924_12445 -NGPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt40_*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	ID	Description
1	69.5	11.5	1188	1 TIRK_ECOLI	P08956 escherichia
2	69.5	11.5	1427	1 REST_HUMAN	P30622 homo sapien
3	69	11.4	353	1 YKFL_YEAST	P35735 saccharomyc
4	68	11.2	378	1 O30A_DROME	Q9vle5 drosophila
5	65.5	10.8	1353	1 XDH_CALVI	P08793 calliphora
6	65	10.7	211	1 G12_ANOGA	Q17040 anopheles g
7	65	10.7	320	1 RLPA_RICPR	Q92del rickettsia
8	65	10.7	534	1 FM02_CAVPO	P36366 cavia porce
9	65	10.7	620	1 MUTL_TREPA	O83325 treponema p
10	65	10.7	680	1 GAG_SCVLA	P32503 saccharomyc
11	64.5	10.7	286	1 THHB_HAEIN	P45134 haemophilus
12	64.5	10.6	937	1 YHML_YEAST	P38856 saccharomyc
13	64	10.6	619	1 GLK3_HUMAN	Q13093 homo sapien
14	63.5	10.5	223	1 Y838_METJA	Q58248 methanococc
15	63.5	10.5	345	1 YDE9_SCHPO	Q10442 schizosacch
16	63.5	10.8	502	1 N02C_MESVI	Q9mud6 mesothicqma
17	63.5	10.5	519	1 LNT_SYNY3	P74055 synchocyst
18	63.5	10.5	1083	1 MAN1_YEAS	P22855 saccharomyc

19	63	10.4	419	1 P47K_PSECL	P31521 pseudomonas
20	63	10.4	423	1 P47A_CANBO	P21245 candida boi
21	63	10.4	465	1 Y065_MYCPN	P75612 mycoplasma
22	62	10.2	419	1 P47B_CANBO	Q00319 candida boi
23	62	10.2	601	1 PEPE_LACLA	Q9cev7 lactococcus
24	61.5	10.2	120	1 ARD2_ECOLI	P52148 escherichia
25	61.5	10.2	153	1 ML1A_XENLA	P51048 xenopus lae
26	61.5	10.2	452	1 NORM_BACSU	Q31855 bacillus su
27	61.5	10.2	819	1 AKIH_SERMA	P27725 serratia ma
28	61.5	10.5	1687	1 YDHA_SCHPO	Q92355 schizosacch
29	61.5	10.2	1901	1 YCF1_TOBAC	P12222 nicotiana t
30	61	10.1	111	1 IMMI_ECOLI	P08701 escherichia
31	61	10.1	176	1 TBAP_HUMAN	Q01658 homo sapien
32	61	10.1	292	1 NLA_DROME	Q9x218 drosophila
33	61	10.1	656	1 YC26_PORPU	P51392 porphyra pu
34	61	10.1	731	1 MAS2_MYCLE	O32913 mycobacteri
35	61	10.1	919	1 GLK3_RAT	P42264 rattus norv
36	60.5	10.3	354	1 MHHS_WOLSU	P31884 wolfinella s
37	60.5	10.0	516	1 YWCA_BACSU	P39599 bacillus su
38	60.5	10.0	547	1 NU5M_AGSU	P24884 musculus suu
39	60.5	10.0	913	1 CAD4_MOUSE	P39038 mus musculu
40	60	9.9	309	1 META_YERPE	Q8zar4 yersinia pe
41	60	9.9	329	1 RIR2_BACSU	P50621 bacillus su
42	60	9.9	354	1 DP42_RHIME	Q92xb8 rhizobium m
43	60	9.9	356	1 Y359_AQUAE	O66685 aquifex aco
44	60	9.9	388	1 TRBL_PYRAB	Q9v188 pyrococcus
45	60	9.9	440	1 CYAD_BORPE	P11091 bordetella

ALIGNMENTS

RESULT 1

ID	TIRK_ECOLI	STANDARD;	PRT; 1188 AA.
AC	P08956;		
DT	01-NOV-1988 (Rel. 09, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Type I restriction enzyme EcoKI R protein (EC 3.1.21.3) (R.EcoKI).		
GN	HSR OR HSR OR B4350.		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12; PubMed=3323532;		
RX	MEDLINE=88118919; PubMed=3323532;		
RA	Loenen W.A.M., Daniel A.S., Braymer H.D., Murray N.E.;		
RT	"Organization and sequence of the hsd genes of Escherichia coli		
RT	K-12.";		
RL	J. Mol. Biol. 198;159-170(1987).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MG1655;		
RX	MEDLINE=95334382; PubMed=7610040;		
RA	Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,		
RT	Blattner F.R.;		
RT	"Analysis of the Escherichia coli genome VI: DNA sequence of the		
RT	region from 92.8 through 100 minutes.";		
RL	Nucleic Acids Res. 23:2105-2119(1995).		
RN	[3]		
RP	SEQUENCE OF 1-27 FROM N.A.		
RC	STRAIN=K12 / C563;		
RX	MEDLINE=91317743; PubMed=1650347;		
RA	Waite-Rees P.A., Keating C.J., Moran L.S., Slatko B.E., Hornstra L.J.,		
RT	Benner J.S.;		
RT	"Characterization and expression of the Escherichia coli Mr		
RT	restriction system.";		
RL	J. Bacteriol. 173:5207-5219(1991).		
CC	!- FUNCTION: THE ECKI ENZYME RECOGNIZES 5'AACN(6)GTC-3'. SUBUNIT R		
CC	IS REQUIRED FOR BOTH NUCLEASE AND ATPASE ACTIVITIES, BUT NOT FOR		
CC	MODIFICATION.		

CC CC CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give random
 CC double-stranded fragments with terminal 5'-phosphate; ATP is
 CC simultaneously hydrolysed.
 CC -1- SUBUNIT: THE TYPE I RESTRICTION/MODIFICATION SYSTEM IS COMPOSED
 CC OF THREE POLYPEPTIDES R, M AND S.
 CC -1- MISCELLANEOUS: TYPE I RESTRICTION AND MODIFICATION ENZYMES ARE
 CC COMPLEX, MULTIFUNCTIONAL SYSTEMS WHICH REQUIRE ATP, S-ADENOSYL
 CC METHIONINE AND MG(2+) AS CO-FACTORS AND, IN ADDITION TO THEIR
 CC ENDONUCLEOLYTIC AND METHYLASE ACTIVITIES, ARE POTENT DNA-DEPENDENT
 CC ATPASES.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1084
 CC ONWARD AND IS SHORTER (1090 AA) DUE TO A FRAMESHIFT.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06545; CAA29791.1; ALT_FRAME.
 DR EMBL; U14003; AAA97247.1; .
 DR EMBL; AE000505; AAC77306.1; .
 DR EMBL; X54198; CAA38116.1; ALT_INIT.
 DR PIR; Q00648; NDECKR.
 DR REBASE; 980; EcoKI.
 DR Ecogene; EGI0459; hsdR.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF00271; helicase_C; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 KW Restriction system: Hydrolase; DNA-binding; ATP-binding;
 KW Complete proteome.
 FT DNA_BIND 449 488 H-T-H MOTIF (BY SIMILARITY).
 FT NP_BIND 490 496 ATP (BY SIMILARITY).
 FT CONFLICT 629 696 DAVKIALTPALHTVQIFGEPPVRYRTRAVIDGLIDOD
 FT PPIQITRNAQEGVLSGEQVERISP -> ECGKNRSHH
 FT PGATYCADFRAGLPLPYGRYRFRSDRGAYSDDHPOR
 FT AGGSLQRRAGRAH (IN REF. 1).
 FT
 SQ SEQUENCE 1188 AA; 136100 MW; CC0423F2A435E578 CRC64;

Alignment Scores:
 Pred. No.: 4,31 Length: 1188
 Score: 69,50 Matches: 30
 Percent Similarity: 37,82% Conservative: 15
 Best local Similarity: 25,21% Mismatches: 33
 Query Match: 11,49% Indels: 41
 DB: 1 Gaps: 5

US-09-856-221-4 (1-332) x TLRK_ECOLI (1-1188)

QY 21 AAAAATATCTGGAACCCAA-----CAGGCACAACTCAGGCAGCTGCCCTTC 71
 ||| ||||| ||| ::||| ||||| ||||| ::|||
 Db 169 LysGlnGlnLeuGlnGlnValArgGluIysAlaGlnThrGlnAlaGluValGluAla 188
 QY 72 CTACAAACAAATTCACGAATACAGCGTGTGATAACTGGCTA----- 113
 ||| ||| ||||| ||||| ::|||
 Db 189 GlnGlnGlnLysLeu-----ValAlaLeuasnGlyIleAlaIleLeuGluGlyLys 206
 QY 114 -----CCTGGCGATTGGCGGCTATATTATTCAGTTTAT 149
 ::| ||||| ||||| ::|||
 Db 207 GlnGlnGluThrGluAlaGlnThrGlnAlaArgLeuAlaLeuGluAlaGlnLeuAla 226
 QY 150 GACTTGGCTGTTCCCTGTTGTGATGGCTGAACAA-----ACTTACCAGTATGAA 200
 ::| ||| ||||| ||||| ::|||
 Db 227 GlnLysAsnAlaGluLeuAlaLysGlnThrGlnGluGlnArgLysAlaIleHisLysGlu 246
 QY 201 TTGAACGATAAAGTGCTACGCTTCATTAAAGCCCGTGGCTGGCATGCCTATGCTGTG 260
 ::| ||||| ||||| ::|||
 Db 247 IleThrAspGlnAlaIleLys----- 253

QY 261 TTGTTACAGCTGAAACCTTGATGCTGAATTTGCACAGATGCAAAAAAACTATTTC 317
 ||||| ||||| ::|||
 Db 254 -----ArgThrLeuasnLeuSerGluGluGluSerArgPheLeu 266
 RESULT 2
 REST_HUMAN STANDARD; PRT; 1427 AA.
 AC P30622;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
 DE Sternberg intermediate filament associated protein).
 GN RSN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Peripheral blood monocytes;
 RX MEDLINE=92289675; PubMed=1600942;
 RA Bilbe G., Delabie J., Bruegggen J., Richener H., Asselbergs F.A.M.,
 RA Cerletti N., Sorg C., Odink K., Tarsay L., Wiesendanger W.,
 RA de Wolf-Peters C., Shipman R.;
 *Restin: a novel intermediate filament-associated protein highly
 RT expressed in the Reed-Sternberg cells of Hodgkin's disease.*;
 RL EMBO J. 11:2103-2113(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=92405160; PubMed=1356075;
 RX Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
 CLIP-170 links endocytic vesicles to microtubules.;
 RL Cell 70:887-900(1992).
 CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
 CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
 CC CYTOSKELETON.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
 CC OF HODGKIN'S DISEASE.
 CC -1- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X64838; CAA46050.1; .
 DR EMBL; M97501; AAA35693.1; .
 DR PIR; S22695; S22695.
 DR Genew; HGNC:10461; RSN.
 DR MIM; 179838; .
 DR InterPro; IPR000938; CAP-Gly.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF01302; CAP_GLY; 2.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS00845; CAP_GLY_1; 2.
 DR PROSITE; PS0245; CAP_GLY_2; 2.
 DR Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
 FT DOMAIN 78 120 CAP-GLY 1.
 FT DOMAIN 143 204 SER-RICH.
 FT DOMAIN 232 274 CAP-GLY 2.
 FT DOMAIN 304 331 SER-RICH.
 FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
 FT DOMAIN 1408 1421 CCHC-BOX.
 FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1069 1069 D -> E (IN REF. 2).
 SQ SEQUENCE 1427 AA; 160989 MW; 0A4F16DD94254E8 CRC64;

US-09-856-221-4 (1-332) x O30A_DROME (1-378)

QY	45	GCACAACTCAGCCACAGCTGGCTTCCTACAAAGCAAAATTCAGCAATACACCGTTGTAT	104
Db	279	AlaGlnThrValIleValIleAlaTyrMetValMetIlePheAlaAsnSerValValLeu	298
QY	105	ACTGGCTACGTGGCGGATGGCGCTATTATTATCATGTTTATCATCTGGCTGCTTCC	164
Db	299	TyrTyrValAlaAsnGluLeu-----TyrPheGlnSerPheAspIleAlaIleAla	315
QY	165	CTGTGTTTGATGCTGCAACAACTTACCAGTAGTAATTGAAC-----	206
Db	316	-----AlaTyrGluSerAsnTrpMetAspPheAspVal	326
QY	207	-----GATAAAGCTGTACGTTTCATT-----	245
Db	327	AspThrGlnLysThrLeuLysPheLeuIleMetArgSerGlnLysProLeuAla-----	344
QY	246	GGCACTTATGCTGGTTTGTGTAGCAGGTGAAACCTGTAGCTGAAATTGGCACAGATGGAA	305
Db	345	-----IleLeuValICLVGLVThrTyrPrometAsnLeuLysMetLeuGln	359

RESULT 5		
XDH_CALVI	STANDARD	PRT: 1353 AA.
ID	XDH_CALVI	
AC	P08793;	
CD	01-NOV-1988 (Rel. 09, Created)	
DT	01-NOV-1988 (Rel. 09, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	de Xanthine dehydrogenase (EC 1.1.1.204) (XD).	
GN	XDH.	
OS	Calliphora vicina (Blue blowfly) (Calliphora erythrocephala).	
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;	
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
OC	Muscomorpha; Oestroidea; Calliphoridae; Calliphora.	
OX	NCBI_TaxID=7373;	
RN	[1]	
RP	SEQUENCE FROM N. A.	
RX	MEDLINE=90185213; PubMed=2516831;	
RA	Houde M., Tiverton M.C., Brogeger F.;	
RT	"Divergence of the nucleotide sequences encoding xanthine	
RT	dehydrogenase in Calliphora vicina and Drosophila melanogaster.";	

dehydrogenase in Calliphora vicina and Drosophila melanogaster." ;
 gene 85:391-402(1989).
 [2]
 SEQUENCE OF 208-367 FROM N.A.
 MEDLINE=88137956; PubMed=2830167;
 Rocher-Chambonnet C., Berreuer P., Houde M., Tiverson M.C.,
 Lepesant J.-A., Bregegere F.;
 "Cloning and partial characterization of the xanthine dehydrogenase
 gene of Calliphora vicina, a distant relative of Drosophila
 melanogaster";
 gene 59:201-212(1987).
 CC -1- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O -> urate + NADH.
 CC -1- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Peroxisomal.
 CC -1- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.
 CC -1- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```
CC -----(see http://www.iab.org/about/iab-announcements)-----
```

```
DR ENBL; X07229; CAA30189.1; -.
```

```
DR ENBL; X07323; CAA30281.1; -.
```

```
DR ENBL; X07324; CAA30281.1; JOINED.
```

```
DR ENBL; X07325; CAA30281.1; JOINED.
```

```
DR ENBL; M18423; AAA27879.1; -.
```



```

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "the genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria.";
CC -!- SIMILARITY: BELONGS TO THE RLPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235271; CAA14847.1; -.
DR InterPro; IPR005132; Lipoprotein_13.
DR Pfam; PF03330; Lipoprotein_13; 1.
DR TIGRPFAMs; TIGR00413; rlpa; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 320 AA; 37353 MW; DBB80B1EF68B834A CRC64;

Alignment Scores:
Pred. No.: 13.5 Length: 320
Score: 65.00 Matches: 25
Percent Similarity: 45.35% Conservative: 14
Best Local Similarity: 29.07% Mismatches: 27
Query Match: 10.74% Indels: 20
DB: 1 Gaps: 5

US-09-856-221-4 (1-332) x RLPA_RICPR (1-320)
QY 12 GTATTGCAAAAACAACTATCTGGAACCAACAGCAGCAAACTCAGGCAGTGGCCCTC 71
Db 2 llelleMetArgAsnTyrlsPhelysLeuGluProLeuLysGlnCysGlnGluAlaLaphe 21
QY 72 CTCAAGCAAAATTCACCAATACAGCTGTAT -----AACTGGCTA 113
Db 22 LysArgTyrlsValHisSerThrAlaLeuTyrllelleAlaLeuGlnAlaAsnIrpSer 41
QY 114 CQTGGGGGATTCGGCGCTATTTATTATCATGCTTTTATGACTTGGCTGTTCCCTGTTGTTG 173
Db 42 Arg---ArgLeu-----lleTyrlsLeuSerLeuTyrlsPheCysIys 54
QY 174 ATGGCTGNA -----CAAACTTACCAATGATTAATGAACAT -----AAA 212
Db 55 lleLysAsnCysTyrlsThrPheLysPheGluGlnSerAsnArgPheillelleThrLys 74
QY 213 GCTGTACGCTTCATTAAG 230
Db 75 AlaGluArgllelleLys 80

RESULT 8
FM02_CAVPO STANDARD; PRT; 534 AA.
AC P36366; Q05194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylaniline monooxygenase [N-oxide forming] 2 (EC 1.14.13.8)
DE (pulmonary flavin-containing monooxygenase 2) (FM0 2) (Dimethylaniline
DE oxidase 2) (FM0 1B).
DN FM02 OR FM0-2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hartley; TISSUE=Lung;
RX MEDLINE=9306545; PubMed=1306120;
RA Nikbakht K.N., Lawton M.P., Philpot R.M.;
RT "Guinea pig or rabbit lung flavin-containing monooxygenases with
RT distinct mobilities in SDS-PAGE are allelic variants that differ at

```

```

RT only two positions.";
RL Pharmacogenetics 2:207-216(1992).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF
CC A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES.
CC -!- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) = N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD AND MAGNESIUM ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Mitosomal.
CC -!- TISSUE SPECIFICITY: LUNG.
CC -!- POLYMORPHISM: THERE ARE TWO ALLELIC FORMS (A AND B).
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10037; AAB59631.1; -.
DR InterPro; IPR000960; Flav_cont_mnoxgn.
DR Pfam; PF00743; FMO-like; 1.
DR PRINTS; PR00370; FMOXYGENASE.
KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family; Acetylation; Polymorphism; Magnesium.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 190 195 NADP (POTENTIAL).
FT NP_BIND 521 521 V -> L (IN B FORM).
FT VARIANT 529 529 S -> F (IN B FORM).
SQ SEQUENCE 534 AA; 60810 MW; 22C6A0A06EB59955 CRC64;

Alignment Scores:
Pred. No.: 13.9 Length: 534
Score: 65.00 Matches: 21
Percent Similarity: 37.89% Conservative: 15
Best Local Similarity: 22.11% Mismatches: 37
Query Match: 10.74% Indels: 22
DB: 1 Gaps: 2

US-09-856-221-4 (1-332) x FM02_CAVPO (1-534)
QY 3 GAAGCGCGAGTATTCACAAAACAACTATCTGGAACCAACAGCAGCAAACTCAGGCACAG 62
Db 423 GlnSerGlnlleValGlnThrAsnTyrlsValasp ----- 433
QY 63 CTGGCCTTCCTACAAAGCAAAATTCAGCAATACAGCTTGTATTAAGTGGCTACGTGGCGCA 122
Db 434 -----TyrLeuAspGluLeuAlaLeuGluilleGlyAlaLysProAsp 447
QY 123 TTGGCGGCTATTATTATCATGCTTTTATGACTTGGCTGTTCCCTGTTGTTGATGGCTGAA 182
Db 448 LeuileSerPheLeuLeuLysaspProGluLeuAlaValLysLeuCysPheGlyProCys 467
QY 183 CAAACTTACCACTATGAATTCAGCATAAAGCTGTAGCTTCATTAAAGCCGGTGGCTGG 242
Db 468 AsnSerTyrlsGlnTyrls-----ArgLeuValGlyProGlyGlnItrp 480
QY 243 CATGGCACCTTATGCTGTTGTTAGCAGGTGAACACCTTCATGCTG 287
Db 481 GluGlyAlaArgAlaAlaIleLeuThrGlnLysGlnArgIleLeu 495

RESULT 9
MUTL_TREPA STANDARD; PRT; 620 AA.
ID MUTL_TREPA
AC O83325;
DT 15-DEC-1998 (Rel. 37, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR TP0303.

```

```

OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RA spirochete.";
RT Science 281:375-388(1998).
RL
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MTHL/HEXB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5001210; AAC65291.1; ALT_INIT.
CC HSSP; P23367; IBKN.
CC TIGR; TP0303; -.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR002099; DNA_mis_repair.
CC Pfam; PF01119; DNA_mis_repair; 1.
CC Pfam; PF02518; HATPase_C; 1.
CC TIGRFAMs; TIGR00585; mutL; 1.
CC PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
CC DNA repair; Complete proteome.
CC SEQUENCE 620 AA; 68495 MW; 0604AD802671EA36 CRC64;

Alignment Scores:
Pred. No.: 14.1 Length: 620
Score: 65.00 Matches: 25
Percent Similarity: 33.96% Conservative: 11
Best Local Similarity: 23.58% Mismatches: 38
Query Match: 11.07% Indels: 32
DB: 1 Gaps: 5

US-09-856-221-4 (1-332) x MUTL_TREPA (1-620)
QY 321 TTTCACAAATAGTTTTTTTCCA----- 301
Db 294 PheProAsnGlyThrPheProValAlaCysLeuPheLeuThrValAsnSerGluArgIle 313
QY 300 -----TCGTGCGCAAAATTCAGCATCAAGGTTTCACCTGCTCAACAAACACAGCATAAAGTGC 247
Db 314 AspPheAsnIleHisProAlaLysLysGluValHisLeu-----GlnAspTyrAla 330
QY 246 CATGCCAGCGCCGGCTTAATGACGGTA---CAGCTTTATCGTTCAATTCATCTAGTGT 190
Db 331 HisIleArgHisThrLeuSerArgSerValAlaHisPheTyrArgGlnCysThrIleAla 350
QY 189 AAGTTTGTTCAG-----CCATCAACACAGGGAACCAACCAAGTCAT 148
Db 351 HisTyrValArgAlaGluProAlaHisAlaProAlaThrGlnGlyAsnAlaProThrHis 370
QY 147 AAAACT-----GATAATAAATAGCCCAATCCGCGCCAGTAGCC 109
DB:

```

```

Db 371 SerSerProCysThrGlyValArgGluProAlaAlaProCysAlaHisThrPro 390
QY 108 AGTTATACACGGCTGTAT 91
Db 391 ArgTyrGluSerLeuPhe 396
RESULT 10
GAG_SCVLA STANDARD; PRT; 680 AA.
ID GAG_SCVLA
AC P32503;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Major coat protein.
GN GAG.
OS Saccharomyces cerevisiae virus L-A (scv-L-A).
OC Viruses; dsRNA viruses; Totiviridae; Totivirus.
OX NCBI_TaxID=11008;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214077; PubMed=2651431;
RA Icho I., Wickner R.B.;
RT "The double-stranded RNA genome of yeast virus L-A encodes its own
RT putative RNA polymerase by fusing two open reading frames.";
RL J. Biol. Chem. 264:6716-6723(1989).
CC -!- PTM: ACETYLATION IS NECESSARY FOR VIRAL ASSEMBLY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J04692; AAA50506.1; -.
CC PIR; S26764; S26764.
CC Coat protein; acetylation.
CC MOD_RES 1 1 ACETYLATION (BY YEAST MAK3).
CC SEQUENCE 680 AA; 75994 MW; 32353DB87F05C943 CRC64;

Alignment Scores:
Pred. No.: 14.2 Length: 680
Score: 65.00 Matches: 21
Percent Similarity: 41.30% Conservative: 17
Best Local Similarity: 22.83% Mismatches: 36
Query Match: 10.74% Indels: 18
DB: 1 Gaps: 2

US-09-856-221-4 (1-332) x GAG_SCVLA (1-680)
QY 51 ACTCAGCACAGCTGGCTTCTCTCAAAAGCAAATTCAGCAATACAGGTTGTAACTGG 110
Db 269 SerLysValMetLeuSerAlaLeuArgLysTyrValAsnHisAsnArgLeuTyrAsn--- 287
QY 111 CTACGTGGCGATGGCGCTATTTATTATCATGTTTATCATCTTGCTGCTTCCCTGTGT 170
Db 288 -----GlnPheTyrThrAlaAlaGlnLeuLeuAla 297
QY 171 TTGATG-----GCTCAACAAACTTACCAGTATCAATTTGAACGAT 209
Db 298 GlnIleMetMetLysProValProAsnCysAlaGlyTyrAlaThrLeuMetHisasp 317
QY 210 AAAGCTGTACGCTTCATTAAAGCCCGGTCCTGGCATGGCATTATGCTGTTGTTAGCA 269
Db 318 AlaLeuValAsnIleProLysPheGlySerIleArgGlyArgTyrProPheLeuLeuSer 337
QY 270 GGTGAACCTTGATGCTGAATTTGGCACAGATGCAA 305
Db 338 GlyAspAlaLeuIleGlnAlaThrAlaLeuGlu 349
RESULT 11
TEHB_HAEIN

```

[illegible]

```

Db 253 ValileysHis1leThrThrlsLeu---ValArgSerLeuGluGluHisLeuIleGlu 271
QY 321 AAAGATGAA 329
Db 272 AspAspLys 274
RESULT 13
GLK3_HUMAN STANDARD; PRT; 919 AA.
AC Q13003; Q16136; Q13004.
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Glutamate receptor, ionotropic kainate 3 precursor (Glutamate receptor
DE 7) (GluR-7) (GluR7) (Excitatory amino acid receptor 5) (EAA5).
GN GRK3 OR GLUR7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS SER-310 AND ARG-352.
RC TISSUE=Fetal brain;
RX MEDLINE=95236038; PubMed=7719709;
RA Nutt S.L., Hoo K.H., Rampersad V., Deverill R.M., Elliott C.E.,
RA Fletcher E.J., Adams S.-L., Korczak B., Foides R.L., Kamboj R.K.;
RT "Molecular characterization of the human EAA5 (GluR7) receptor: a
RT high-affinity kainate receptor with novel potential RNA editing
RT sites.";
RL Recept. Channels 2:315-326(1994).
RN [2]
RP SEQUENCE OF 268-320 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94174381; PubMed=8128318;
RA Puranam R.S., Eubanks J.H., Heinemann S.F., McNamara J.O.;
RT "Chromosomal localization of gene for human glutamate receptor
RT subunit-7.";
RL Somat. Cell Mol. Genet. 19:581-588(1993).
RN [3]
RP VARIANT SER-310.
RX MEDLINE=20574843; PubMed=11124978;
RA Schiffer H.H., Swanson G.T., Maaliyah E., Heinemann S.F.;
RT "Unequal expression of allelic kainate receptor GluR7 mRNAs in human
RT brains.";
RL J. Neurosci. 20:9025-9033(2000).
CC -1- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
CC MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE
CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. THIS RECEPTOR BINDS
CC DOMOATE > KAINATE >> L-GLUTAMATE - QUISSOLATE >> AMPA - NMDA.
CC -1- SUBUNIT: HOMOMERIC. CAN ALSO COASSEMBLE WITH EITHER GRK4 OR
CC GRK5 TO FORM HETEROMERIC RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U16127; AAB60407.1; -
CC EMBL; U16128; AAC50421.1; -
CC EMBL; S69349; AAB30157.1; -
CC HSP; P19431; IGR2.
CC Genew; HGNC:4581; GRK3.
CC MIM; 138243; -
CC InterPro; IPR001828; ANF_receptor.
CC InterPro; IPR001320; Ion_glu_receptor.
CC InterPro; IPR001311; SBP_glu_receptor.

```

```

DR Pfam; PF00060; liq_chan; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR ProDom; PD000500; Ion_glu_receptor; 1.
DR SMART; SM00079; PBPe; 1.
KW Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
KW Transmembrane; RNA editing; Polymorphism.
FT SIGNAL 1 31
FT CHAIN 32 919
FT DOMAIN 32 563
FT TRANSMEM 564 584
FT TRANSMEM 637 657
FT TRANSMEM 821 841
FT CARBOHYD 70 70
FT CARBOHYD 76 76
FT CARBOHYD 278 278
FT CARBOHYD 381 381
FT CARBOHYD 415 415
FT CARBOHYD 426 426
FT CARBOHYD 433 433
FT CARBOHYD 548 548
FT VARIANT 310 310
FT VARIANT 352 352
FT VARIANT 303 303
FT CONFLICT 303 303
FT SEQUENCE 919 AA; 104031 MW; 2CD7E49B00195E68 CRC64;
Alignment Scores:
Pred. No.: 18.9 Length: 919
Score: 64.00 Matches: 22
Percent Similarity: 40.18% Conservative: 23
Best Local Similarity: 19.64% Mismatches: 61
Query Match: 10.58% Indels: 6
DB: 1 Gaps: 1
US-09-856-221-4 (1-332) x GLK3_HUMAN (1-919)
QY 6 GCGGCAGTATTGCAAAAACCTATCTGGAACCCCAACAGCAGCAAACTCAGGCACAGCTG 65
Db 293 SerAlaIleValGluLysTrpSerMetGluArgLeuGlnAlaAlaProArgAlaGluSer 312
QY 66 GCCTCTCTACAAACCAANTTCAGCAATACAGCGTTGTATACCTGGCTACGTGGCGGCGATTG 125
Db 313 GlyLeuLeuAspGlyValMetMetThrAspAlaAlaLeuLeuTyAspAlaValHisIle 332
QY 126 GCGGCTATTATTATCATCAGTTTTATGACTTGGCTGTTCCCTGTTGTTGATGGCTGAACAA 185
Db 333 ValSerValCysTyrGlnArgAlaProGlnMetThrValAsnSerLeuGlnCysHisGln 352
QY 186 ACTTACCAAGTATGAATTGAACGATAAAGCTGTACGCTTCAATTAAGCCCGCTGGCGCAT 245
Db 353 HisTysAlaTrpArgPheGlyArgPheMetAsnPheIleLysGluAlaGlnTrpGlu 372
QY 246 GGCACCTATTGCTGTTTGTAGCAGGTGAACCC-----TTGATGCTG 287
Db 373 GlyLeuThrGlyArgIleValPheAsnLysThrSerGlyLeuArgThrAspPheAspLeu 392
QY 288 AATTTGGCAGATGAGAAAAAACTATTGGAAAAA 323
Db 393 AspIleIleSerLeuLysGluAspGlyLeuGluLys 404
RESULT 14
Y838_METJA STANDARD; PRT; 223 AA.
AC Q58248;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0838.
GN MJ0838.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.

```

```

OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts C.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii."; Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO M.JANNASCHII MJ0575.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67528; AAB98843.1; -.
DR TIGR; MJ0838; -.
DR InterPro; IPR003745; DUF166.
DR Pfam; PF02593; DUF166; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 223 AA; 25837 MW; 1F592F76C84EC37F CRC64;

Alignment Scores:
Pred. No.: 19,9 Length: 223
Score: 63.50 Matches: 16
Percent Similarity: 56.41% Conservative: 6
Best Local Similarity: 41.03% Mismatches: 10
Query Match: 10.50% Indels: 7
DB: 1 Gaps: 2

US-09-856-221-4 (1-332) x Y838_METJA (1-223)
QY 147 TATGACTGTGCTGTCCTCGTTTGATGGCTGAACAACTTACCAGTAT----- 197
Db 52 TyrAspLeuPheIleThrThrLeuAsnProAspLeuThrThrGluLeuValArgLys 71
QY 198 -----GAATTGAACGATAAAGCTAGCTTCAATTAAGCCGGTGGCGTGGCAGGC 248
Db 72 IleLysGluLeuAsnLysAla-----PheValLeuValGlyAlaTrpLysGly 88

RESULT 15
YDE9_SCHPO
ID YDE9_SCHPO STANDARD; PRT; 345 AA.
AC Q10442;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative mitochondrial carrier C12B10.09.
GN SPAC12B10.09
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Rger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Calbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Certutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussey D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (Potential).
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; 270721; CAA94699.1; -.
DR InterPro; IPR001993; Mitoch_carrier.
DR Pfam; PF00153; mito_carr; 3.
DR PROSITE; PS00215; MITOCH_CARRIER; 2.
KW Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
KW Transmembrane; Transport.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 319 339 POTENTIAL.
SQ SEQUENCE 345 AA; 38451 MW; 39081A6AAC984B2C CRC64;

Alignment Scores:
Pred. No.: 20,4 Length: 345
Score: 63.50 Matches: 28
Percent Similarity: 46.84% Conservative: 9
Best Local Similarity: 35.44% Mismatches: 29
Query Match: 10.50% Indels: 13
DB: 1 Gaps: 4

US-09-856-221-4 (1-332) x YDE9_SCHPO (1-345)
QY 40 AACAGGCACAACTCAGG-----CACAGCTGGCCTCTCTACAAAGCAATTACGCA 90
Db 61 AsnArgHisGlnLeuLeuArgThrCysHisPheThrProSer---LysArgHisSerAla 79
QY 91 ATACAGCGTGTGTATACCTAGCTACGGCGGATTTGGCGGCTATTATTATATCAGTTTTATG 150
Db 80 Met--SerPhePheGluAlaLeuGlyAlaGlyIleCysAla-----G 93
QY 151 ACTTGCGCTGTTCCCTGCTGTTGTTGCTGCTGAACAACTTACCAGTATGAATGAACGATA 210
Db 93 lylleuAlaValAspLeuSerLeuPheProIleAspThrLeuLysThrArgLeuGlnAla 113
QY 211 AAGCTGTACGCTTTCATTAAAGCCGGTGGCTGGCATGGCAGCTTATGCTGGTTG 263
```

Db 113 ysGly--GlyPheValLysAsnGlyGlyPheHisGlyValTyrArgGlyLeu 129

Search completed: January 15, 2003, 15:46:35
Job time : 11.8532 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:44:13 ; Search time 31.2773 Seconds

(Without alignments)
4374.264 Million cell updates/sec

Title: US-09-856-221-4
Perfect score: 605
Sequence: 1 gtgaagcgcagctatgca.....atttggaagaaagatgaacgg 332

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlip
-MODEL=frame-n2p.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09856221/runat_15012003_153925_12457/app_query_fasta_1.1948
-DB=SPTREMBL_21 -QFMT=FASTA -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09856221@cgn1_1.171.8runat_15012003_153925_12457 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Length	ID	Description
1	534	88.3	2523	2 Q93RP0 Q93rp0 xenorhabdus

2	446	73.7	2516	2 Q9RN43	Q9rn43 photorhabdu
3	418	69.1	2504	2 Q85160	Q85160 photorhabdu
4	395.5	65.4	2376	2 Q9F923	Q9f923 serratia en
5	373	61.7	2538	2 Q93RW7	Q93rn7 xenorhabdus
6	268	44.3	1565	2 Q85156	Q85156 photorhabdu
7	219	36.2	1279	2 Q93RP4	Q93rp4 xenorhabdus
8	179.5	29.7	1189	2 Q85152	Q85152 photorhabdu
9	76	12.6	121	2 P96782	P96782 xiphophilus
10	75	12.4	168	7 Q8SNH2	Q8snh2 xiphophilus
11	74.5	12.3	598	2 Q91815	Q91815 treponema p
12	73	12.1	439	17 Q58431	Q58431 pyrococcus
13	72.5	12.0	1606	3 Q9UW87	Q9uw87 candida alb
14	72	11.9	269	16 Q69492	Q69492 mycobacteri
15	71	11.7	332	2 Q9F7M4	Q9f7m4 uncultured
16	70.5	11.7	553	3 Q8X0K4	Q8x0k4 neosporos
17	70.5	11.7	596	2 Q9K2N7	Q9k2n7 treponema p
18	70.5	11.7	598	16 Q88138	Q88138 treponema p
19	70.5	11.7	1309	16 Q9KS77	Q9ks77 vibrio chol
20	70	11.6	725	4 Q96MR5	Q96mr5 homo sapien
21	69.5	11.5	427	2 Q8VUK1	Q8vuk1 paracoccus
22	69.5	11.5	467	16 Q97RH0	Q97rh0 neisseria m
23	69.5	11.5	601	16 Q99253	Q99253 streptococ
24	68.5	11.3	258	16 Q55454	Q55454 synecocyst
25	68.5	11.3	312	5 Q61892	Q61892 caenorhabdi
26	68.5	11.3	363	17 Q971K7	Q971k7 sulfolobus
27	68	11.2	503	16 Q8ZGI8	Q8zgi8 yersinia pe
28	68	11.2	582	5 Q9N515	Q9n515 caenorhabdi
29	68	11.2	698	16 Q990J8	Q990j8 staphylococ
30	68	11.2	1056	3 Q9P8P8	Q9p8p8 candida alb
31	67.5	11.2	120	2 Q9KJ14	Q9kji4 klebsiella
32	67.5	11.5	335	8 Q03370	Q03370 procavia ca
33	67.5	11.2	596	2 Q9L814	Q9l814 treponema p
34	67.5	11.2	610	5 Q20798	Q20798 caenorhabdi
35	67.5	11.2	680	12 Q87024	Q87024 saccharomyc
36	67.5	11.2	703	2 Q9R714	Q9r714 helicobacte
37	67.5	11.2	703	16 Q9ZJX3	Q9zjx3 helicobacte
38	67.5	11.2	707	2 Q51811	Q51811 helicobacte
39	67.5	11.2	716	5 Q9N8E5	Q9n8e5 trypanosoma
40	67.5	11.2	741	2 Q52269	Q52269 helicobacte
41	67.5	11.2	744	16 Q9ZK22	Q9zk22 helicobacte
42	67	11.1	105	2 P96780	P96780 haemophilus
43	67	11.4	335	8 Q9Y32	Q9y32 orycteropus
44	67	11.4	379	8 Q9THD6	Q9thd6 orycteropus
45	67	11.1	590	10 Q9SLW2	Q9slw2 olea europae

ALIGNMENTS

RESULT 1				
Q93RP0	Q93RP0	PRELIMINARY;	PRT;	2523 AA.
AC	Q93RP0;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	XptAI protein.			
GN	XPTAI.			
OS	Xenorhabdus nematophilus.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Xenorhabdus.			
OX	NCBI_TaxID=628;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PMF1296;			
RA	MEDLINE=21218513; PubMed=11319082;			
RX	Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;			
RT	"Sequence analysis of insecticidal genes from Xenorhabdus nematophilus			
RT	PMF1296."			
EL	Appl. Environ. Microbiol. 67:2062-2069 (2001).			
DR	EMBL: AJ308438; CAC38401.1;			
DR	InterPro: IPR000953; Chromo.			
DR	SMART: SM00298; CHROMO; I.			
SQ	SEQUENCE 2523 AA: 286999 MW: 3159852E0655B5B1 CRC64;			

Alignment Scores:
 Pred. No.: 1,1e-55 Length: 2523
 Score: 534.00 Matches: 103
 Percent Similarity: 96.36% Conservative: 3
 Best Local Similarity: 93.64% Mismatches: 4
 Query Match: 88.26% Indels: 0
 DB: 2 Gaps: 0

US-09-856-221-4 (1-332) x Q93RP0 (1-2523)

Qy 3 GAAGCGCGAGTATTCGAAACAACTATCTGGAACCCACAGGACAACTCAGGCACAG 62
 Db 2234 GluAlaAlaValLeuGlnLysAsnTyrLeuGluThrGlnGlnAlaGln 2253

Qy 63 CTGGCGCTTCTACAAAGCAATTCAGCAATACAGCGTTGTATTAAGTGGCTGGCGGA 122
 Db 2254 LeuAlaPheLeuGlnSerLysPheSerAsnAlaAlaLeuTyrAsnTrpLeuArgGlyArg 2273

Qy 123 TTGGCGCTATTTATATCAAGTTTATGACTTGGCTGTTCCCTGTTGATGGCTGAA 182
 Db 2274 LeuSerAlaIleTyrGlnPheTyrAspLeuAlaValSerLeuCysLeuMetAlaGlu 2293

Qy 183 CAAACTTACAGTATGAAGCAATGAAGCTGTACCTTCATTAAAGCCGCTGCTGG 242
 Db 2294 GlnThrTyrGlnTyrGluLeuAsnAsnAlaAlaAlaHisPheIleLysProGlyAlaTrp 2313

Qy 243 CATGGCACTATGCTGGTTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGATG 302
 Db 2314 HisGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuAsnLeuAlaGlnMet 2333

Qy 303 GAAAAAATCTTTGGAAGAAAGATGAACGG 332
 Db 2334 GluLysSerTyrLeuGluLysAspGluArg 2343

RESULT 2
 Q9RN43 PRELIMINARY; PRT; 2516 AA.

ID Q9RN43
 AC Q9RN43; 2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Toxin A (Toxin complex protein).
 GN TCDA.
 OS Photobacterium luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Photobacterium.
 CX NCBI_TaxID=29488;
 RN SEQUENCE FROM N.A.
 RC STRAIN-W-14;
 RA Bowen D.J., Wegrich L.M., Roberts J.L., Petell J.K.,
 RT "Expression of Photobacterium luminescens tcda gene in maize confers
 corn rootworm tolerance."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC STRAIN-W14;
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
 RL MEDLINE=21185117; PubMed=11286884;
 RT "The tc genes of Photobacterium: a growing family."
 RL Trends Microbiol. 9:185-191(2001).
 DR EMBL; AF188483; AAF05542.1; -;
 DR EMBL; AF346500; AAL18486.1; -;
 DR InterPro; IPR001589; Actbind actin.
 DR PROSITE; PS00019; ACTININ.1; UNKNOWN.1.
 SQ SEQUENCE 2516 AA; 282952 MW; 8250A0650B614B99 CRC64;

Alignment Scores:
 Pred. No.: 4.74e-45 Length: 2516
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12

Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 2 Gaps: 0

US-09-856-221-4 (1-332) x Q9RN43 (1-2516)

Qy 3 GAAGCGCGAGTATTCGAAACAACTATCTGGAACCCACAGGACAACTCAGGCACAG 62
 Db 2218 GluAlaAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2237

Qy 63 CTGGCGCTTCTACAAAGCAATTCAGCAATACAGCGTTGTATTAAGTGGCTGGCGGA 122
 Db 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2257

Qy 123 TTGGCGCTATTTATATCAAGTTTATGACTTGGCTGTTCCCTGTTGATGGCTGAA 182
 Db 2258 LeuAlaAlaIleTyrPheGlnPheTyrAspLeuAlaValAlaArgCysLeuMetAlaGlu 2277

Qy 183 CAAACTTACAGTATGAAGCAATGAAGCTGTACCTTCATTAAAGCCGCTGCTGG 242
 Db 2278 GlnAlaTyrArgTrpGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 2297

Qy 243 CATGGCACTATGCTGGTTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGATG 302
 Db 2298 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317

Qy 303 GAAAAAATCTTTGGAAGAAAGATGAACGG 332
 Db 2318 GluAspAlaHisLeuLysArgAspLysArg 2327

RESULT 3
 O85160 PRELIMINARY; PRT; 2504 AA.

ID O85160
 AC O85160;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Insecticidal toxin complex protein tcba.
 GN TCBA.
 OS Photobacterium luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Photobacterium.
 CX NCBI_TaxID=29488;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-W-14;
 RA Bowen D., Nocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
 RT "Insecticidal toxins from the bacterium Photobacterium luminescens."
 RL Science 280:2129-2132(1998).
 RN [2]
 RC STRAIN-W14;
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
 RL MEDLINE=21185117; PubMed=11286884;
 RT "The tc genes of Photobacterium: a growing family."
 RL Trends Microbiol. 9:185-191(2001).
 DR EMBL; AF047457; AAC38627.1; -;
 DR EMBL; AF346498; AAL18460.1; -;
 DR InterPro; IPR000566; Lipoclin_cytfabp.
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN.1.
 SQ SEQUENCE 2504 AA; 280654 MW; 8F29B1693D1047CE CRC64;

Alignment Scores:
 Pred. No.: 1.15e-41 Length: 2504
 Score: 418.00 Matches: 78
 Percent Similarity: 84.55% Conservative: 15
 Best Local Similarity: 70.91% Mismatches: 17
 Query Match: 69.09% Indels: 0
 DB: 2 Gaps: 0

```

US-09-856-221-4 (1-332) x 085160 (1-2504)
QY 3 GAAGCGCAGTATTCGAAACCAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 2212 GluAlaAlaGluMetGlnLysGluThrLeuLysThrGlnGlnAlaGlnAlaGln 2231
QY 63 CTGGCCCTTCCTCAACAGCAATTCAGCAATACAGCTTGTATTAACCTGGCTACGTGGCGA 122
DB 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 2251
QY 123 TTGGCGGCTATTATATACAGTTTATGAGCTTGTGCTGTTCCCTGTTGTTGTTGCTGAA 182
DB 2252 LeuSerGlyIleThrPheGlnPheThrAspLeuAlaValSerArgCysLeuMetAlaGln 2271
QY 183 CAACCTACAGTATGAATGAGCAATAGCTGTACGCTTCATTAGCCGGTGGCTGG 242
DB 2272 GlnSerTyrGlnTrpGluAlaAsnAspSerIleSerPheValLysProGlyAlaTrp 2291
QY 243 CATGGCCTATCTGCTGTTGTAGCAGGTGAAACCTTGTATGCTGAATTTGGCACAGATG 302
DB 2292 GlnGlyThrTrpAlaGlyLeuLysCysGlyGluAlaLeuIleGlnAsnLeuAlaGlnMet 2311
QY 303 GAAAAAATCTTTGAAAAAGATCAACGG 332
DB 2312 GluGluAlaTrpLeuLysTrpGluSerArg 2321

RESULT 4
QYF923 ID QYF923 PRELIMINARY; PRT: 2376 AA.
AC QYF923;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DE "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus
DE SepA.
GN SepA.
OS Serratia entomophila.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Plasmid pADAP.
OC Serratia.
OX NCBI_TaxID=42906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ALM02;
RX MEDLINE=20416224; PubMed=10960097;
RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
RT "Plasmid-located pathogenicity determinants of Serratia entomophila,
RT the causal agent of amber disease of grass grub, show similarity to
RT the insecticidal toxins of Photobacterium luminescens.";
RL J. Bacteriol. 182:5127-5138(2000).
DR EMBL; AF135182; AAG09642.1; -.
KW Plasmid.
SQ SEQUENCE 2376 AA; 262649 MW; E418DACE22DBB7BF CRC64;

Alignment Scores:
Pred. NO.: 5,99e-39 Length: 2376
Score: 395.50 Matches: 80
Percent Similarity: 82.73% Conservative: 11
Best Local Similarity: 72.73% Mismatches: 18
Query Match: 65.37% Indels: 1
DB: 2 Gaps: 1

US-09-856-221-4 (1-332) x QYF923 (1-2376)
QY 3 GAAGCGCAGTATTCGAAACCAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 2087 GluGlyAlaGluLeuGlnLysThrTrpLeuGluThrGlnGlnAlaGlnAlaGln 2106
QY 63 CTGGCCCTTCCTCAACAGCAATTCAGCAATACAGCTTGTATTAACCTGGCTACGTGGCGA 122
DB 2107 LeuAlaPheLeuGlnSerLysPheAsnThrAlaLeuTyrSerTrpLeuArgGlyArg 2126
QY 123 TTGGCGGCTATTATATACAGTTTATGAGCTTGTGCTGTTCCCTGTTGTTGCTGAA 182

US-09-856-221-4 (1-332) x Q93RN7 (1-2538)
QY 3 GAAGCGCAGTATTCGAAACCAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 2249 GluAlaAlaGlnMetGlnValGluThrGlnGlnAlaHisThrGlnAlaGln 2268
QY 63 CTGGCCCTTCCTCAACAGCAATTCAGCAATACAGCTTGTATTAACCTGGCTACGTGGCGA 122
DB 2269 LeuGluLeuLeuGlnArgLysPheThrAsnLysAlaLeuTyrSerTrpMetArgGlyLys 2288
QY 123 TTGGCGGCTATTATATACAGTTTATGAGCTTGTGCTGTTCCCTGTTGTTGCTGAA 182
DB 2289 LeuSerAlaIleThrTrpGlnPheAspLeuThrGlnSerPheCysLeuMetAlaGln 2308
QY 183 CAACCTTACAGTATGAATTCAGCAATAGCTGTACGCTTCATTAGCCGGTGGCTGG 242
DB 2309 GluAlaLeuArgGluLeuThrAspAsnGlyValThrPheIleArgGlyAlaTrp 2328
QY 243 CATGGCCTATCTGCTGTTGTAGCAGGTGAAACCTTGTATGCTGAATTTGGCACAGATG 302
DB 2329 AsnGlyThrAlaGlyLeuMetAlaGlyThrLeuLeuLeuAsnLeuAlaGlnMet 2348
QY 303 GAAAAAATCTTTGAAAAAGATCAACGG 332
DB 2349 GluLysValTrpLeuGluArgAspGluArg 2358

RESULT 6

```

```

085156
ID 085156 PRELIMINARY; PRT; 1565 AA.
AC 085156;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Insecticidal toxin complex protein Tceb.
GN TCEB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., French-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";
RL Science 280:2129-2132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA French-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF047028; AAC38629.1; -;
DR EMBL; AF346499; AAL18472.1; -;
SQ SEQUENCE 1565 AA; 175717 MW; 66C3AB96C5FA8397 CRC64;

Alignment Scores:
Pred. No.: 1,5e-23 Length: 1565
Score: 268.00 Matches: 52
Percent Similarity: 61.82% Conservative: 16
Best Local Similarity: 47.27% Mismatches: 42
Query Match: 44.30% Indels: 0
DB: 2 Gaps: 0

US-09-856-221-4 (1-332) x 085156 (1-1565)
QY 3 GAAGCGCAGTATTGCAAAAAAATCTCTGGAACCCAGCAGCAAACTCAGGCACAG 62
Db 1254 LysAlaAlaGlnThrSerLeuGlnGlnAlaLysAlaGlnGlnValGlnIleArgThrMet 1273
QY 63 CTGGCTCTCTACAAAGCAAAATACAGCAGTGTGTATACTGGCTAGCGGGCGA 122
Db 1274 LeuThrThrLeuThrThrArgPheThrGlnAlaThrLeuThrGlnTrpLeuSerGlyGln 1293
QY 123 TTGGCGGCTATTATTATCATGTTTATCATGTTGGCTGCTTTCCTGCTGTTGATGGCTGAA 182
Db 1294 LeuSerAlaLeuThrGlnAlaThrGlnAlaThrAspAlaValValAlaLeuCysLeuSerAlaGln 1313
QY 183 CAAACTTACCATGATGAATGAACGATTAAGCTGTAGCTTCAATTAAGCCGCTGCTGG 242
Db 1314 AlaCysTrpGlnTrpGlnLeuGlyAspTrpAlaThrThrPheIleGlnThrGlyThrTrp 1333
QY 243 CATGCCACTTATCTGCTGGTTGTTACACAGCTGAACCTTGATGCTGAATTTGGCACAGATG 302
Db 1334 AsnAspHisThrArgGlyLeuGlnValGlyLeuGlnValGlyLeuGlnValGlnHisGlnMet 1353
QY 303 GAAAAAACTATTGGAAAAAGATGAACGG 332
Db 1354 GluAlaAlaThrLeuValArgHisGluArg 1363

RESULT 7
ID 093RP4 PRELIMINARY; PRT; 1279 AA.
AC 093RP4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)

```

```

DE XptD1 protein (Fragment).
GN XPTD1.
OS Xenorhabdus nematophilus.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Xenorhabdus.
OX NCBI_TaxID=628;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMF1296;
RX MEDLINE=21218513; PubMed=11319082;
RA Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus
RT PMF1296.";
RL Appl. Environ. Microbiol. 67:2062-2069(2001).
DR EMBL; AJ308438; CAC38397.1; -;
DR InterPro; IPR001993; MitochCarrier.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1279 AA; 146105 MW; 11841387EF906B98 CRC64;

Alignment Scores:
Pred. No.: 1.24e-17 Length: 1279
Score: 219.00 Matches: 41
Percent Similarity: 61.39% Conservative: 21
Best Local Similarity: 40.59% Mismatches: 39
Query Match: 36.20% Indels: 0
DB: 2 Gaps: 0

US-09-856-221-4 (1-332) x 093RP4 (1-1279)
QY 30 CTGGAACCCAGCAGCAAACTCAGGCACAGCTGCCTCTCTACAAAGCAATTCAGC 89
Db 980 LeuGluAlaGlnHisGlnAspGlnValLeuLeuGluThrTyrSerAsnArgPheThr 999
QY 90 AATACAGCGTTGTATAACCTACGTGGCGGATTTGGCGCTATTATTATCATGTTTAT 149
Db 1000 AsnAspAlaLeuThrMetTrpMetIleSerGlnIleSerGlyLeuGlnAlaTyr 1019
QY 150 GACTTGCTGTTTCCCTGCTGTTGATGGCTGACAACTTACCATGATGATTCAGCAT 209
Db 1020 AspAlaValAsnSerLeuCysLeuLeuAlaGlnAlaSerTrpGlnTrpGlnThrGlyGln 1039
QY 210 AAAGCTGTAGCTTTCATTAAAGCCGGTGGCTGGCATGGCACTTATGCTGGTTGTAGCA 269
Db 1040 TyrAspMetAsnPheValGlnSerGlyLeuTrpAsnAspLeuTrpGlnGlyLeuVal 1059
QY 270 GGTGAACCTTGATGCTGAATTTGGCAGACATGAAAAAACTATTGAAAAAGATGAA 329
Db 1060 GlyGluHisLeuLysLeuAlaLeuGlnArgMetAspGlnAlaTyrLeuGlnHisAsnThr 1079
QY 330 CGG 332
Db 1080 Arg 1080

RESULT 8
ID 085152 PRELIMINARY; PRT; 1189 AA.
AC 085152;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Insecticidal toxin complex protein Tcab.
GN TCAB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., French-Constant R.H.;

```

```

RT "Insecticidal toxins from the bacterium Photobacterium luminescens.",
RL Science 280:2129-2132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA French-Constant R.H.;
RT "The toxins of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF046867; AAC38624.1; -.
DR EMBL; AF346497; AAL18450.1; -.
DR InterPro; IPR002927; Virus_HS.
DR Pfam; PF01550; Virus_HS; 1.
SQ SEQUENCE 1189 AA; 131466 MW; 7DAA3AE0B8B8127 CRC64;

Alignment Scores:
Pred. No.: 7,34e-13 Length: 1189
Score: 179.50 Matches: 37
Percent Similarity: 58.16% Conservative: 20
Best Local Similarity: 37.76% Mismatches: 40
Query Match: 23.67% Indels: 1
DB: 2 Gaps: 1

US-09-856-221-4 (1-332) x 085152 (1-1189)
QY 33 GAAACCCACAGCACAACCTCAGCAGCAGCTGGCTTCTCTACAAAGCAATTCAGCAAT 92
ID Q8SNH2 PRELIMINARY; PRT; 121 AA.
DB 921 GluThrGluGlnAlaAsnAlaGlnAlaIleTyrAspLeuGlnThrThrArgPheThrGly 940
QY 93 ACAGCGTTGTATACGTAGCTGGCGGATGGCGCTATTATATCATCATTTATGAC 152
ID Q8SNH2;
DB 941 GlnAlaLeuTyrAsnTrpMetAlaGlyArgLeuSerAlaLeuTyrTyrGlnMetTyrAsp 960
QY 153 TTGCGTGTCCCTGTTGTGATGCTGACAAACTTACCATGATGAAATTCAGCAATAA 212
ID Q8SNH2;
DB 961 SerThrLeuProIleCysLeuGlnAlaAlaLeuValGlnGluGlyGluLys 980
QY 213 --GCTGTAGCTTCAATTAAGCCCGCTGCTGGCGATGCGCTTATGCTGGTTTATGCA 269
ID Q8SNH2;
DB 981 GluSerAspSerLeuPheGlnValProValTyrAsnAspLeuTyrGlnGlyLeuAla 1000
QY 270 GGTGAACCTTGATGCTGAATTTGCACAGATGCAAAACATTTTCGAAAAA 323
ID Q8SNH2;
DB 1001 GlyGluGlyLeuSerGluLeuGlnLysLeuAspAlaIleTyrPheAlaArg 1018

RESULT 9
P96782
ID P96782 PRELIMINARY; PRT; 121 AA.
AC P96782;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major outer membrane protein P2-type 7 (Fragment).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OC NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith-Vaughan H.C., Sriprakash K.S., Mathews J.D., Kemp D.J.;
RT "Non-encapsulated Haemophilus influenzae in Aboriginal infants with
RT otitis media: Prolonged carriage of P2 porin variants and evidence for
RT horizontal P2 gene transfer.";
RL Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89638; ABA49486.1; -.
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
FT NON_TER 1
FT NON_TER 121
SQ SEQUENCE 121 AA; 13482 MW; D0C7B580FC7F9766 CRC64;

Alignment Scores:
Pred. No.: 2.82 Length: 168
Score: 75.00 Matches: 16
Percent Similarity: 52.31% Conservative: 18
Best Local Similarity: 24.62% Mismatches: 29
Query Match: 12.40% Indels: 2
DB: 7 Gaps: 2

US-09-856-221-4 (1-332) x Q8SNH2 (1-168)
QY 54 CAGCAGCAGCTGGCTTCTCTACAAAGCAAAATTCAGC---AATACAGCTGTATTAACGG 110
ID Q8SNH2;
DB 23 GluValGlnIleSerTyrCysAspSerArgThrAsnGluAsnIleProLysGlnAsnTrp 42
QY 111 CTAGCTGGCGGATGGCGGCTATTATATCATGTTTATGACTGGCTGTTCCCTGTGT 170
ID Q8SNH2;
DB 43 Met---AsnLysValSerSerGluTyrProAspTyrTyrLysGluGluThrGluThrCys 61
QY 171 TTGATGCTGCAACAACTTACAGTATGATGAATTCAGCATAAAGCTGTACGCTTCATTAAG 230
ID Q8SNH2;
DB 109 LysAlaLysHisGluLysSerTyrPheValSerProGly 121

RESULT 10
Q8SNH2
ID Q8SNH2 PRELIMINARY; PRT; 168 AA.
AC Q8SNH2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MHC class I alpha antigen (Fragment).
OS Xiphophorus hellerii guentheri.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OC NCBI_taxid=188717;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=XIHGB11;
RX MEDLINE=21655117; PubMed=11797104;
RA Figueroa F, Mayer W.E., Sato A., Zaleska-Rutczynska Z., Hess B.,
RA Tichy H., Klein J.;
RT "MHC class I genes of swordtail fishes, Xiphophorus: variation in the
RT number of loci and existence of ancient gene families.";
RL Immunogenetics 53:695-708(2001).
DR EMBL; AF411653; AAM18399.1; -.
FT NON_TER 1
FT NON_TER 168
SQ SEQUENCE 168 AA; 19734 MW; 5545EC2B83F8009 CRC64;

Alignment Scores:
Pred. No.: 2.82 Length: 168
Score: 75.00 Matches: 16
Percent Similarity: 52.31% Conservative: 18
Best Local Similarity: 24.62% Mismatches: 29
Query Match: 12.40% Indels: 2
DB: 7 Gaps: 2

US-09-856-221-4 (1-332) x Q8SNH2 (1-168)
QY 54 CAGCAGCAGCTGGCTTCTCTACAAAGCAAAATTCAGC---AATACAGCTGTATTAACGG 110
ID Q8SNH2;
DB 23 GluValGlnIleSerTyrCysAspSerArgThrAsnGluAsnIleProLysGlnAsnTrp 42
QY 111 CTAGCTGGCGGATGGCGGCTATTATATCATGTTTATGACTGGCTGTTCCCTGTGT 170
ID Q8SNH2;
DB 43 Met---AsnLysValSerSerGluTyrProAspTyrTyrLysGluGluThrGluThrCys 61
QY 171 TTGATGCTGCAACAACTTACAGTATGATGAATTCAGCATAAAGCTGTACGCTTCATTAAG 230
ID Q8SNH2;
DB 109 LysAlaLysHisGluLysSerTyrPheValSerProGly 121

```

Db 62 LeuValLysGlnGlnThrPheLysTyrAsnLeuGluLeuAlaLysThrArgPheAsnGln 81

QY 231 CCCGTCGCTGGCAT 245

Db 82 ThrGlyGlyValHis 86

RESULT 11

Q9L815

ID Q9L815 PRELIMINARY: PRT: 598 AA.

AC Q9L815;

DT 01-OCT-2000 (T-EMBLrel. 15, Created)

DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)

DE TprD3.

DE TprD3.

GN TPRD.

OS Treponema pallidum (subsp. pertenue) (Yaws treponeme).

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI_TaxID=168;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-GAUTHIER;

RX MEDLINE=20200377; PubMed=10735882;

RA Centurion-Lara A., Sun E.S., Barrett L.K., Castro C., Lukehart S.A.,

RA Van Voorhis W.C.;

RA "Multiple alleles of treponema pallidum repeat gene D in treponema

RT pallidum isolates."

RL J. Bacteriol. 182:2332-2335(2000).

RL EMBL: AF217538; AF64130.1; -;

DR InterPro; IPR000847; HTH_LysR.

DR InterPro; IPR003857; MOSP_C.

DR InterPro; IPR003857; MOSP_Nterm.

DR Pfam; PF02722; MOSP_C; 1.

DR Pfam; PF02707; MOSP_N; 1.

DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.

SQ SEQUENCE 598 AA; 64562 MW; 60D1BE78DOA32278 CRC64;

Alignment Scores:

Pred. No.: 3 46 Length: 598

Score: 74.50 Matches: 36

Percent Similarity: 38.13% Conservative: 17

Best Local Similarity: 25.90% Mismatches: 40

Query Match: 12.31% Indels: 46

DB: 2 Gaps: 5

US-09-856-221-4 (1-332) x Q9L815 (1-598)

QY 53 TCAGCAGCAGCTGCTTCTCTACAAAGCAAAATTCAGCAATACAGCGTTGTATACTG-- 110

Db 19 SerGlyTyrAlaGlyValLeuThrProGlnVal-SerGlyThrAlaGlnLeuGlnTrpGl 38

QY 111 -----CTACGTGGCGATTGGCGCTATTATTATCA 142

Db 38 yIleAlaPheGlnLysAsnProHisThrValProGlyLysHisThrHisGlyPheArgTh 58

QY 143 GTTTTATGACTGCTGCTTCTCTGTTGATGGCTGAACAACTAC----- 191

Db 58 rThrAsnSerLeuThrIleSerLeuProLeuValSerLysHisThrHisThrArgAtGgl 78

QY 192 -----CAGTATGAATTGAACGATAAGCTGTACGC----- 221

Db 78 yGluAlaArgSerGlyValTrpAlaGlnLeuLysAspLeuAlaValGluLeuAl 98

QY 222 -----TTCAATTAAGCCGGTGC----- 239

Db 98 aSerSerLysSerSerThrAlaLeuSerPheThrLysProThrAlaSerPheGlnAlaTh 118

QY 240 -----TGGCATGGCACTTATGCTGTTGTAGCAGGTGAACCTTGATGCTGAATTT 292

Db 118 rLeuHisCysTyrGlyAlaTyrLeuThrValGlyThrSerProSerCysValValAsnPh 138

QY 293 GCACAGATGMAAA-----AACTNTTGGAAAAGATGAACGG 332

Db 138 eAlaGlnLeuTrpLysProPheValThrArgAlaTyrSerGluLysAspThrArg 156

RESULT 12

O58431

ID O58431 PRELIMINARY: PRT: 439 AA.

AC O58431;

DT 01-AUG-1998 (T-EMBLrel. 07, Created)

DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)

DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)

DE Hypothetical protein PH0700.

GN PH0700.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku F.,

RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,

RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

RA Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."

RL DNA Res. 5:55-76(1998).

DR EMBL: AP000003; BAA25791.1; -;

KW Hypothetical protein, Complete proteome.

SQ SEQUENCE 439 AA; 48361 MW; EB9F404553B15362 CRC64;

Alignment Scores:

Pred. No.: 5 17 Length: 439

Score: 73.00 Matches: 18

Percent Similarity: 55.56% Conservative: 12

Best Local Similarity: 33.33% Mismatches: 20

Query Match: 12.07% Indels: 4

DB: 17 Gaps: 2

US-09-856-221-4 (1-332) x O58431 (1-439)

QY 132 ATTATTATCAGTTTAT-----GACTGTGCTGTTTCCCTGTGTTGATGCTGAA 182

Db 122 IlePheTyrMetPhePheLeuAlaAlaThrIleProLeuSerLeuLeuSerArg 141

QY 183 CAACACTTACCAGTATGATTAAGCTGATGCTTCAATTAAGCCGCGCTGCTG 242

Db 142 GlupPheArgTyrGluLeuAsnLysAlaIleGlyArgPheAsnGluLeuGlyGlyTip 161

QY 243 CATGGCATTATCTGTTGTTAGCAGGTGAACCTTGATG 284

Db 162 ---GlyTrpValValGlyLeuLeuGlyPheThrLeuLeu 174

RESULT 13

Q9UW87

ID Q9UW87 PRELIMINARY: PRT: 1606 AA.

AC Q9UW87;

DT 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)

DE MRP-like transporter.

GN MRP1.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5476;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV SS;

RX MEDLINE=21927371; PubMed=11929516;

RA Theiss S., Kreschmar M., Nichterlein T., Hof H., Agabian N.,

RA Hacker J., Koehler G.A.;

RT "Functional analysis of a vacuolar ABC transporter in wild-type


```
Db 211 ArgAspAlaValAsnSerAlaSerLeuGlyLysSerSerLysSerSerTyrGlnMet 230
QY 195 -----TATGAATTGAACGATAAAGCT 215
Db 231 SerProLysAsnIleAsnGluAlaLeuHisGluValAlaMetAspIleAsnGluGlyAla 250
QY 216 ---GTACGCTTCATTAAAGCCGCGTGCCTGCCAT----- 245
Db 251 AspIleValMetValLysProGlyMetProTyrLeuAspIleIleSerLysValLysGlu 270
QY 246 -----GGCACTTATGCTGGTTTGTACCAAGTCAACCTTGATGCTGAATTG 293
Db 271 ThrPheLysValProThrPheAlaTyrGlnValSerGlyGluTyrSerMetLeuLysLeu 290
QY 294 GCACAGATGGAAAAAAGCTATTGGAAAAAGAT 326
Db 291 Ala---IleAspLysGlyTrpLeuGluSerAsp 300
```

Search completed: January 15, 2003, 15:54:41
Job time : 39.2773 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 10.6207 Seconds
(without alignments)
3506.895 Million cell updates/sec

Title: US-09-856-221-1

Perfect score: 793

Sequence: 1 tgcgcagcactcacttat.....gtgacccaactgatacagtt 449

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+nt2p.model -DEV-xlp
-Q/cg21/USPTO_spoel/US09856221/runat_15012003_153924_12445/app_query.fasta_1.1948
-DB-Swissprot_40 -QPMT-fascan -SUFFIX-rsp -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE-pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-US09856221 -CGN_1_1_38_@runat_15012003_153924_12445 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	9.8	468	1 ALS2_CANAL	Q74657 candida alb
2	76	9.8	640	1 UL06_HSV7J	P52455 human herpe
3	76	9.6	318	1 TF2D_MESAU	P53360 mesocricetu
4	76	9.6	2555	1 P5S3_BACSU	P39847 bacillus su
5	76	9.6	2799	1 EDD_HUMAN	O95071 homo sapien
6	75	9.5	316	1 TF2D_MOUSE	P29037 mus musculu
7	75	9.5	1080	1 HDA4_CHICK	P83038 gallus gall
8	73.5	9.3	297	1 TF2D_XENLA	P27633 xenopus lae
9	73.5	9.3	339	1 TF2D_HUMAN	P20226 homo sapien
10	73.5	9.3	622	1 PRNB_MYCLE	P54744 mycobacteri
11	73.5	9.3	1718	1 RPO_SHVX	Q04575 shallot vir
12	73	9.2	256	1 YJJP_ECOLI	P39402 escherichia
13	72.5	9.1	805	1 E2F_DROME	Q27368 drosophila
14	72.5	9.1	2688	1 ZEP1_MOUSE	Q03172 mus musculu
15	72	9.1	442	1 ODO2_RAT	Q01205 rattus norv
16	72	9.1	488	1 CWF8_SCHPO	Q04011 schizosacch
17	72	9.1	881	1 GLND_VIBCH	Q9KPV0 vibrio chol
18	72	9.1	1145	1 PR22_YEAST	P24384 saccharomyc

19	71.5	9.0	201	1 RR4_SPIOLO	P13788 spinacia ol
20	71.5	9.0	470	1 RAN1_SCHPO	P08092 schizosacch
21	71	9.0	300	1 TF2D_TRIFL	Q92117 trimeresaur
22	71	9.0	302	1 TF2D_TRIGA	Q92146 trimeresaur
23	71	9.0	509	1 YRM4_CAEEL	Q10051 caenorhabdi
24	71	9.0	886	1 YFIQ_ECOLI	P76594 escherichia
25	71	9.0	2441	1 CBP_MOUSE	P45481 mus musculu
c	70	9.0	481	1 MURC_TREPA	O83361 treponema p
27	70	8.8	614	1 TULL_SCHPO	Q09715 schizosacch
28	70	8.8	624	1 PPS1_CAVPO	O54820 c bifunctio
29	70	8.8	2303	1 POLG_TMEVG	P08545 t genome po
30	69.5	8.8	368	1 ROX1_YEAST	P25042 saccharomyc
31	69.5	8.8	373	1 CIS2_MYCTU	Q10529 mycobacteri
c	69.5	8.8	601	1 NCPR_SALTR	P19618 saimo trutt
33	69.5	8.8	2832	1 NDVB_RHIME	P20471 rhizobium m
34	69	8.7	201	1 CTF1_HUMAN	Q16619 homo sapien
35	69	8.7	302	1 TF2D_CHICK	O13270 gallus gall
36	69	8.7	381	1 ASG1_YEAST	P38986 saccharomyc
37	69	8.7	423	1 MTB1_BACAM	P23941 bacillus am
38	69	8.7	2303	1 POLG_TMEVB	P08544 t genome po
c	68.5	8.8	396	1 T5P1_CAEEL	Q17848 caenorhabdi
40	68.5	8.6	1260	1 ALS1_CANAL	P46590 candida alb
41	68.5	8.6	1419	1 ALAL_CANAL	Q13368 candida alb
c	68.5	8.8	1659	1 VIT_ONCMY	Q92093 oncorhynch
43	68	8.6	129	1 RL7_CHLMU	P38001 chlamydia m
44	68	8.6	447	1 DNAA_SYNY3	P49995 synechocyst
45	68	8.6	448	1 YAP1_CHICK	P46936 gallus gall

ALIGNMENTS

RESULT 1	ALS2_CANAL	STANDARD;	PRT;	468 AA.
ID	074657;			
AC	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DE	Agglutinin-like protein 2 precursor (Fragment).			
DE	Agglutinin-like protein 2 precursor (Fragment).			
GN	ALS2.			
OS	Candida albicans (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxID=5476;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1161;			
RX	MEDLINE=96440424; PubMed=9765564;			
RA	Hoyer L.L., Payne T.L., Hecht J.E.;			
RT	"Identification of Candida albicans ALS2 and ALS4 and localization of			
RT	als proteins to the fungal cell surface.";			
RL	J. Bacteriol. 180:5334-5343(1998).			
CC	!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.			
CC	!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; AF024580; AAC64235.1;			
DR	Cell adhesion; Glycoprotein; Repeat; Signal.			
KW	SIGNAL			
FT	CHAIN 1 17			
FT	CHAIN 18 >468			
FT	CARBOHYD 253			
FT	CARBOHYD 315			
FT	NON_TER 458			
SQ	SEQUENCE 468 AA: 50203 MW: BFE773EL69ED0FAF CRC64;			

Alignment Scores:

Pred. No.:	2.75	Length:	468
Score:	78.00	Matches:	42
Percent Similarity:	42.18%	Conservative:	20
Best Local Similarity:	28.57%	Mismatches:	53
Query Match:	9.84%	Indels:	32
DB:	1	Gaps:	7

US-09-856-221-1 (1-449) x ALS2_CANAL (1-468)

QY	42	ATTGGTAACGATGGTGCAGACCCCGTTAG	---ACAAAGCGCGCAGTCAACCAATTCG	98
Db	298	ValGlySerSerLeuLusSerLysP	ropheAsnLeuargLeuargGlyTy	317
QY	99	TGATCAATTATCAGCATAAATATGCGGCAACTGGGTCAACGGCGCGGCCTTGGCGGGGAAACG	158	
		:::		
Db	318	GluAlaAsnSerAsnGlypheValIleValAlaThrThrargThrValThrAspSerThr	337	
QY	159	TACTGCAAAATTCGTTAACCGCTTGTTCCTCTCCAGATATAACAAAAAATGCAAACTTA	218	
Db	338	ThrAlaValThrThrLeuProPheAsnProSerValAspLysThrLysThrIleGluLe	357	
QY	219	CTGGCAGCGTTAGCACAAACCCCTATATAACTTACCTACGCTCAATTAATCTGACAATTGATGCTCA	278	
		:::		
Db	358	LeuGlnProIleProThrThrThrIle	-----ThrThr-----	369
		:::		
QY	279	GCCCTTGTCAATACCCATCTATGCGACACCAGCAGATCCGCTCGTACTGCTTACTGCTGCTGC	338	
		:::		
Db	370	TyrValGlyValThr	-----ThrSerTyrSerThrLysThrAla-----	382
		:::		
QY	339	CGTCACCGCTCACAAAGCGGAGGGATGTCCTCGGACAGTAATCGCATGCTACCGCTTT	398	
Db	383	-----ProIleGlyGluThrAlaThrValIleValAspValProTyr	-----	396
QY	399	-----TCGAVTATTCGTA-----	AAATGCCAAGTCGGG	428
Db	397	HisThrThrThrThrValThrSerGluThrThrGlyThrIleThrThrThrThrArg	416	
QY	429	AGTGACCCCACTGATACAGTT	449	
Db	417	ThrAsnProThrAspSerIle	423	

RESULT 2

UL06_HSV17J	STANDARD;	FRQ;	640 AA.
AC	P52455;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Virion protein U76.		
GN	U76.		
OS	Human herpesvirus (type 7 / strain J1) (HHV7).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Alphaherpesvirinae; Simplexvirus.		
ON	NCBI_TaxID=57278;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RA	Nicholas J.;		
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.		
CC	!- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA		
CC	PACKAGING.		
CC	!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,		
CC	EHV-1 56, EBV BRF1, HCMV UL104, AND VZV 54.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstati-		
CC	on the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for comm-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-		
CC	or send an email to license@isb-sib.ch).		
CC			
CC	EMBL; U43400; AAC54737.1; --		
DR	InterPro: IPR002560; Herpes UL6		

DR HSP: P20226; LTGH.
 DR InterPro: IPR000814; TFIID.
 DR Pfam: PF00352; TBP; 2.
 DR PRINTS: PR00686; TIFACTORIID.
 DR PROSITE: PS00351; TFIID; 2.
 KW Transcription regulation: DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 59 72 POLY-GLN.
 FT REPEAT 144 220 1.
 FT REPEAT 234 311 2.
 SQ SEQUENCE 318 AA; 34956 MW; A25FCD27A6422593 CRC64;
 Alignment Scores:
 Pred. No.: 4.28 Length: 318
 Score: 76.00 Matches: 26
 Percent Similarity: 43.27% Conservative: 19
 Best Local Similarity: 25.00% Mismatches: 45
 Query Match: 9.58% Indels: 14
 DB: 1 Gaps: 3
 US-09-856-221-1 (1-449) x TFD2_MESAU (1-318)
 QY 70 TTAGAGAGCCGCCAGTCAACCAATTCGTGATCATATACAGATAAAATGCGGCAACG 129
 Db 52 LeuGluGlu-----GlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 69
 QY 130 CQTCAAGCGCGCCCTGCGCGCGGAAAGCTACTGCAAAATTCGTTAAACCGCTTTGTTCCCT 189
 Db 70 GlnGlnGlnAlaValAlaThrAlaAlaAlaSerValGlnGlnSerThrSer----- 86
 QY 190 CCTCAGATAAACAACAACTGCAAGATTACTGCGCAGACGTTAGCAGCAACGCTATATAC 249
 Db 87 -----GlnGlnSerThrGlnGlyAlaSerGlyGlnThrProGln 99
 QY 250 TTACGTGATATCTGACATGATGGTCAGCGCTG-----TCATTAACCACTATGCG 303
 Db 100 LeupheHisGlnThrLeuThrAlaProteuproglyThrProLeuLeuThrPro 119
 QY 304 ACACGAGAGATCGCTGCTACTGCTAGTGGCTGCGCTGCGCTGCGCTGCGAGGAGG 363
 Db 120 SerProMethThrProMethThrProLeuThrProAlaThrProAlaSerGluSerGly 139
 QY 364 GATTTGCTCGG 375
 Db 140 IleValProGln 143
 RESULT 4
 ID PPS3_BACSU STANDARD; PRT; 2555 AA.
 AC P39847;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Peptide synthetase 3.
 GN PFSC OR PPS3.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=168;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.B., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapius A., Lardinol S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche M., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sorokin A., Tacconi E., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.P., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis"; 249-256(1997).
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE OF 1-859 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95227362; PubMed=7711903;
 RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
 RA Grandi G.;
 RT "A putative new peptide synthase operon in Bacillus subtilis: partial
 characterization";
 RL Microbiology 141:645-648(1995).
 CC -! COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES (BY
 CC SIMILARITY).
 CC -! SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -! SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: Z99113; CAB13715.1; -
 CC EMBL: Z34883; CAA84362.1; -
 CC HSSP: P14687; IAMU.
 DR SubtilList; BG10972; ppsc.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; Condensatn.
 DR InterPro: IPR003880; Peptide attach.
 DR Pfam: PF00503; AMP-binding; 2.
 DR Pfam: PF00550; pp-binding; 2.
 DR Pfam: PF00668; Condensation; 3.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00012; PHOSPHOPANTHETINE; 1.
 DR PROSITE: PS00455; AMP-BINDING; 2.
 DR PROSITE: PS00755; ACYL CARRIER; 2.
 DR Multifunctional enzyme; Ligase; Repeat; Complete proteome.
 KW DOMAIN 972 1039 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2008 2074 ACYL CARRIER (ACP) 2.
 FT BINDING 2038 2038 PHOSPHOPANTHETINE (POTENTIAL).
 SQ SEQUENCE 2555 AA; 287499 MW; 3E50B3395105D5D0 CRC64;
 Alignment Scores:
 Pred. No.: 5.48 Length: 2555
 Score: 76.00 Matches: 43
 Percent Similarity: 34.86% Conservative: 18
 Best Local Similarity: 24.57% Mismatches: 44
 Query Match: 9.58% Indels: 70
 DB: 1 Gaps: 8
 US-09-856-221-1 (1-449) x PPS3_BACSU (1-2555)
 QY 52 GATTTGTCAGAACCCCGCTTTAGAGAGCCGCCAGTCAACCACTTCGT----- 99

```

Db 2234 AspLeuSerAlaGlyTyrGlnGlnAlaAlaGlyGlnThrGlnLeuProProLys 2253
QY 100 ---GATCATATACGATGCGCAACGTCGTCACGCGCGCCCTTCCGCGGAAA 156
Db 2254 ThrAspSerTyrGlnGlnTyrAlaArgGlnGlnTyrAla----- 2268
QY 157 CGTACTGCAATTCGTTACCCGCTTGTTCCTCCACATAAACAAAACATG----- 210
Db 2269 -----GlnSerSerLysLeuIleArgGlu 2276
QY 211 CAAAGTTACTGG-----CAGACGTTAGCAGCAGCGCTATATACTTACCT 255
Db 2277 GluAlaTyrTyrSerValGluGlnGlnAlaGlnLeuProTyrGluIlePro 2296
QY 256 CATATCTGCAATTCAT---GGTCAGCGCTGTCATTCCTCCATCATGCGACAGCAGCA 312
Db 2297 HisHisValAsnIleAspPheSerLysArgAspSerLysSerPheSerLeuThrGluAla 2316
QY 313 GATCCGTCGCTACGCTT----- 330
Db 2317 AspThrAlaValLeuLeuGlnAsnValAsnHisAlaTyrGlyThrAspThrGlnAspIle 2336
QY 331 -----AGTCGTCGCTACCGCTTCCGATTATTCGAAAT 417
Db 2337 LeuLeuThrAlaAlaSerLeuAlaIleCysGluTyrThrGlyGlySerLysLeuArgIle 2356
QY 364 -----GATTTCGCTCGCAGCAGTA 381
Db 2357 AlaMetGluGlyHisGlyArgGluHisIleLeuProGluLeuAspIleSerArgThrVal 2376
QY 382 -----ATGCGGATGACGCTTTCGATTATTCGAAAT 417
Db 2377 GlyTyrPheThrSerMetTyrProAlaLeuIleSerPheGluAsn 2391
RESULT 5
EDD_HUMAN STANDARD; PRT; 2799 AA.
AC 095071: Q9NPL3; 094970;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquitin-protein ligase EDD (EC 6.3.2.-) (Hyperplastic discs
DE protein homolog) (HHYD) (Progesterin induced protein).
GN EDD OR HYD OR KIA00896.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta, and Heart;
RX MEDLINE=99153743; PubMed=10030672;
RA Callaghan M.J., Russell A.J., Woolliatt E., Sutherland G.R.,
RA Sutherland R.I., Watts C.K.W.;
RT Identification of a human HECT family protein with homology to the
RL Drosophila tumor suppressor gene hyperplastic discs.;
RN Oncogene 17:3479-3491(1998).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=21671350; PubMed=11714696;
RA Honda Y., Tojo M., Matsuzaki K., Anan T., Matsumoto M., Ando M.,
RA Saya H., Nakao M.;
RT Cooperation of HECT-domain ubiquitin ligase HHYD and DNA
RT topoisomerase II-binding protein for DNA damage response.;
RN J. Biol. Chem. 277:3599-3605(2002).
RN [3]
RP SEQUENCE OF 1569-2799 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

```

```

RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.04 ANGSTROMS) OF 2391-2455, AND MUTAGENESIS
RP OF CYS-2768
RA MEDLINE=21192643; PubMed=11287654;
RA Deo R.C., Sonenberg N., Burley S.K.;
RT "X-ray structure of the human hyperplastic discs protein: an ortholog
RT of the C-terminal domain of poly(A)-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4414-4419(2001).
CC -1- FUNCTION: E3 ubiquitin-protein ligase which accepts ubiquitin from
CC an E2 ubiquitin-conjugating enzyme in the form of a thioester and
CC then directly transfers the ubiquitin to targeted substrates (By
CC similarity). This protein may be involved in maturation and/or
CC post-transcriptional regulation of mRNA. May play a role in
CC control of cell cycle progression. May have tumor suppressor
CC function. Regulates DNA topoisomerase II binding protein (TopBP1)
CC for the DNA damage response.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Widely expressed. Most abundant in testis and
CC expressed at high levels in brain, pituitary and kidney.
CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-
CC thiolester formation.
CC -1- SIMILARITY: CONTAINS 1 HECT-TYPE E3 UBIQUITIN-PROTEIN LIGASE
CC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licenses@sib-sib.ch).
CC -----
EMBL: AF006010; AAD01259.2;
EMBL: U95000; AAF88143.1;
EMBL: AB020703; BA874919.1;
DB: F0B; 1121; 18-APR-02.
DR InterPro: IPR000569; HECT_domain.
DR InterPro: IPR002004; PABP/HECT.
DR InterPro: IPR000449; UBA_domain.
DR InterPro: IPR003126; Znf_Nrecognin.
DR Pfam: PF00627; UBA; 1.
DR Pfam: PF00632; HECT; 1.
DR Pfam: PF00658; PABP; 1.
DR Pfam: PF02207; zf-UBR1; 1.
DR SMART: SM00119; HECTC; 1.
DR SMART: SM00517; PolyA; 1.
DR SMART: SM00396; Znf-UBR1; 1.
DR PROSITE: PS50237; HECT; 1.
KW Ubiquitin-protein ligase; Nuclear protein; 3D-structure.
FT DOMAIN 2393..2449 PABP-LIKE.
FT DOMAIN 2462..2799 HECT.
FT DOMAIN 1986..2059 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 2036..2059 PRO-RICH.
FT DOMAIN 2329..2348 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 2357..2366 ARG/ASP-RICH (MIXED CHARGE).
FT DOMAIN 2489..2500 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 2737..2757 PRO-RICH.
FT DOMAIN 940..945 POLY-GLU.
FT DOMAIN 980..985 POLY-SER.
FT DOMAIN 1528..1537 POLY-SER.
FT DOMAIN 1671..1681 POLY-SER.
FT DOMAIN 1762..1768 POLY-ALA.
FT BINDING 2768..2768 UBIQUITIN (BY SIMILARITY).
FT MUTAGEN 2768..2768 C->A; LOSS OF UBIQUITIN BINDING.
FT CONFLICT 134..134 E -> P (IN REF. 2).
FT CONFLICT 229..229 S -> K (IN REF. 2).
FT CONFLICT 258..258 S -> Y (IN REF. 2).
FT CONFLICT 374..374 IG -> M (IN REF. 2).
FT CONFLICT 772..772 D -> H (IN REF. 2).

```

```

RL Nucleic Acids Res. 21:2769-2769(1993).
RP [3]
RN SEQUENCE FROM N.A.
RC STRAIN-TW18/TW5; TISSUE=Testis;
RX MEDLINE=97432814; Pubmed=9286694;
RA Trachtulec Z., Hamvas R.M., Forejt J., Lehrach H.R., Vincek V.,
RA Klein J.;
RT "Linkage of TATA-binding protein and proteasome subunit C5 genes in
RT mice and humans reveals synteny conserved between mammals and
RT invertebrates.";
RL Genomics 44:1-7(1997).
CC -!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIID BINDS
CC SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
CC THE POSITION OF TRANSCRIPTION INITIATION.
CC -!- SUBUNIT: BINDS DNA AS A MONOMER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TBP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license@lsb-sib.ch. (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: D01034; BAA00840.1; -
DR EMBL: U63933; AAB53097.1; -
DR PIR: S34437; S34437.
DR HSSP: P20226; ITGH.
DR TRANSFAC; T00796; -
DR MGD; MGI:101838; Tbp.
DR InterPro; IPR000814; TFIID.
DR Pfam; PF00352; TBP; 2.
DR PRINTS; PR00686; TIFACTORIID.
DR PROSITE; PS00351; TFIID; 2.
DR KEGG; 5570; POLY-GLN.
DR DOMAIN 55 70
DR REPEAT 142 218
DR REPEAT 232 309
DR FT VARIANT 70 70 Q -> OQO (IN STRAIN TW18/TW5)..
DR SEQUENCE 316 AA; 34709 MW; A33D776160B4A393 CRC64;
-----
Alignment Scores:
Pred. No.: 5 46 Length: 316
Score: 75.00 Matches: 23
Percent Similarity: 43.16% Conservative: 18
Best Local Similarity: 24.21% Mismatches: 42
Query Match: 9.46% Indels: 12
DB: 1 Gaps: 2
-----
US-09-856-221-1 (1-449) x TF2D_MOUSE (1-316)
QY 97 CGTGATCATTTATCAGCATAAATCGGCAACTCGTCAACGGCGGGCTTCGCGGGGAAA 156
Db 57 ArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaValAlaThrAlaAa 76
QY 157 CGTATCGCAAAATCGGTAAACGGCTTGTCTTCCTCCAGATAAACAAAACATGCAACGT 216
Db 77 AlaSerValGlnGlnSerThrSer-----GlnGln 86
QY 217 TACTGGCAGCGCTTACCAACAGCGCTATATAACTTACGTACATAATCTGCACAAATGATGCT 276
Db 87 ProThrGlnGlyAlaSerGlyInThrProGlnGlnPheHisSerGlnInThrLeuThr 106
QY 277 CAGCGGTG-----TCAPTACCATCTATCGACACACAGCAGATCCGCTACGTCGT 330
Db 107 AlaProLeuProGlyThrThrProLeuTyProSerProMetThrProMetThrProIle 126
QY 331 AGTGTGTCGCTCACCCTCCACAGGGGAGGGGATTCGCTCGG 375
Db 127 ThrProAlaThrProAlaSerGlnSerGlyGlyValProGln 141

```


QY 259 AATCTGACAAATTGATGTCAGCGCTTG-----TCATTACCACTATCGACACAGCA 312
 Db 82 ProGlnThrLeuThrAlaProLeuProGlyAsnThrProLeuThrProSerProLeu 101
 QY 313 GATCCGTCGGTACGTTAGTGTGCGCCGTCACAGCGCGAGGGGATTTGGCT 372
 Db 102 ThrProMetThrProLeuSerProAlaThrProAlaSerGluSerSerGlyIleValPro 121
 QY 373 CGG 375
 Db 122 Gln 122

RESULT 9
 TF2D_HUMAN
 ID TF2D_HUMAN STANDARD; PRT: 339 AA.
 AC P20226; Q16845; Q9UC02;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription initiation factor TFIIID (TATA-box factor) (TATA
 sequence-binding protein) (TFP).
 GN TBP OR TFIIID OR TF2D.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=90302006; PubMed=2363050;
 RA Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;
 RT "Cloning of a transcriptionally active human TATA binding factor.";
 RL Science 248:1646-1649(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90326195; PubMed=2374612;
 RA Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,
 Roeder R.G.;
 RT "Highly conserved core domain and unique N terminus with presumptive
 regulatory motifs in a human TATA factor (TFIID).";
 RL Nature 346:387-390(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Smalley C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP DOMAINS.
 RX MEDLINE=90302006; PubMed=2363050;
 RA Peterson M.G., Tanese N., Pugh B.F., Tijan R.;
 RT "Functional domains and upstream activation properties of cloned
 human TATA binding protein.";
 RL Science 248:1625-1630(1990).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.
 RX MEDLINE=96209823; PubMed=8643494;
 RA Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,
 Burley S.K.;
 RT "Crystal structure of a human TATA box-binding protein/TATA element
 complex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.
 RX MEDLINE=96346176; PubMed=8757291;
 RA Joo Z.S., Chiu T.K., Leiberman P.M., Baikalov I., Berk A.J.,
 Dickerson R.E.;
 RT "How proteins recognize the TATA box.";
 RL J. Mol. Biol. 261:239-254(1996).
 CC -!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
 OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIIID BINDS
 SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
 THE POSITION OF TRANSCRIPTION INITIATION.
 CC -!- SUBUNIT: BINDS DNA AS A MONOMER.

CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE TBP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL; M55654; AAA36731.1; -;
 DR EMBL; M34960; AAC03409.1; -;
 DR EMBL; X54993; CAA38736.1; -;
 DR EMBL; AL031259; CAA20286.1; -;
 DR PIR; A34830; A34830.
 DR PIR; A34831; A34831.
 DR PDB; 1TGH; 01-AUG-96.
 DR PDB; 1CDW; 23-DEC-96.
 DR TRANSFAC; T00794; -;
 DR Genew; HGNC:11588; TBP.
 DR MIM; 600075; -;
 DR InterPro; IPR000814; TFIID.
 DR Pfam; PF00352; TBP; 2.
 DR PRINTS; PR00686; TIFACTORIID.
 DR PROSITE; PS00351; TFIID; 2.
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;
 KW 3D-structure.
 FT DOMAIN 55 95 POLY-GLN.
 FT REPEAT 165 241 1.
 FT REPEAT 255 332 2.
 FT CONFLICT 58 61 MISSING (IN REF. 2).
 FT CONFLICT 187 187 A -> R (IN REF. 1; AAC03409).
 SQ SEQUENCE 339 AA; 37698 MW; A61A578D972B970B CRC64;
 Alignment Scores:
 Pred. No.: 7.93 Length: 339
 Score: 73.50 Matches: 23
 Percent Similarity: 42.72% Conservative: 21
 Best Local Similarity: 22.33% Mismatches: 46
 Query Match: 9.27% Indels: 13
 DB: 1 Gaps: 2
 US-09-856-221-1 (1-449) x TF2D_HUMAN (1-339)
 QY 73 GAAGAAGCCGCCAGTCAAAACCATTCGTGATCATATCAGCATAAAATCGGCAACTCGT 132
 Db 73 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 92
 QY 133 CAACGCCGCCCTTCCCGCGGAAGCTACTGCAAAATTCGTAACCGCTTGTTCCTTCCT 192
 Db 93 GlnGlnGlnAlaValAlaAlaAlaAlaValGlnGlnSerThrSer----- 107
 QY 193 CAGATAACAAAAAAGTCTGAGCAGACGTTAGCAGACGCTAGCAGACGCTATATAACTTA 252
 Db 108 -----GlnGlnAlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 121
 QY 253 CGTCATAATCTGCAATTCGATGGTCAGCCGCTG-----TCATTACCATCTATCGGACA 306
 Db 122 PheHisSerGlnThrLeuThrThrAlaProLeuProGlyThrThrProLeuThrProSer 141
 QY 307 CCAGCAGATCCGTCGCTAGCTGCTTAGTCTGCTGCTCAGCGCTCACAAGCGGAGGGAT 366
 Db 142 ProMetThrProMetThrProLeuThrProAlaThrProAlaThrProAlaSerGluSerGlyIle 161
 QY 367 TTGCCTCGG 375
 Db 162 ValProGln 164
 RESULT 10
 PKNB_MYCLE
 ID PKNB_MYCLE STANDARD; PRT: 622 AA.
 AC P54744;

QY	106	TATCAGCATAAATCGCGCAACGCGCGGCGCTTGCGCGGAAACGT-----	159
Db	225	TyrGlnHis-----ValArgGluAspProIleProSerGlnArgHisGlu	240
QY	160	---ACTGCBAATCGTTAACCGCTTGTTCCTCCTCAGATAACAAAAAAGTCGCAAGT	216
Db	241	GlyIleSerValAspLeuAspAlaValLeuLeuAlaLeuAlaLysAsnProGluAsn	260
QY	217	TACTGCGACGCGTTAGCACACGCGCTATATAACTTACGTCATATCTG-----ACAATT	270
Db	261	ArgTyrGlnThrAlaAlaGlu-----MetArgLaAspLeuLeuLeuArgValHis	276
QY	271	GATGTCAGCGCGTTGTCATTACCATCTATCGGACACACAGAGATCCGTCGCTACTGCTT	330
Db	277	SerGlyGlnProGluAlaProLysValLeuThrAspAlaAspArgSerCysLeuLeu	296
QY	331	AGTGTGCGCGTCACCGCGCTCAAGCGGAGGGGATTGTGCTCGGACA	378
Db	297	SerSerGlyAla-----GlyAsnPheGlyValProArgThr	308
RESULT 11			
RRPO SHVX			
ID	RRPO SHVX	STANDARD;	PRT; 1718 AA.
AC	Q04575;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	RNA replication protein (194 kDa protein) (ORF 1) [Contains: RNA-		
DE	directed RNA polymerase (BC 2.7.7.48); Probable helicase].		
DE	Shallot virus X (ShvX).		
QC	Viruses: ssRNA positive-strand viruses, no DNA stage; Allexivirus.		
OX	NCBI_taxID=31770;		
RN	[1]		
SEQUENCE FROM N.A.			
XX	MEDLINE=93019008; PubMed=1339469;		
XX	Kanyuka K.V., Vishnichenko V.K., Levay K.E., Kondrikov D.Y.,		
XX	Ryabov E.V., Zavrjev S.K.;		
RT	"Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal		
RT	cistron closely related to those of potexviruses and a unique		
RT	arrangement of the 3'-proximal cistrons.";		
PL	J. Gen. Virol. 73:2553-2560(1992).		
CC	-I- FUNCTION: RNA-replication. The central part of this protein		
CC	possibly functions as an ATP-binding helicase.		
CC	-I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +		
CC	{RNA}(N).		
CC	This SWISS-PROT entry is copyright. It is produced through a collabor		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstat		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for comm		
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno		
CC	or send an email to license@isb-sib.ch).		
CC	EMBL; M97264; AAA47787.1; -		
DR	PIR; J01734; J01734.		
DR	InterPro; IPR005123; 2OG-FeII_Oxy.		
DR	InterPro; IPR001788; RNA_dep_RNAPol.2.		
DR	InterPro; IPR000606; Viral_helicase1.		
DR	Pfam; PF00978; RNA_dep_RNAPol2; 1.		
DR	Pfam; PF01443; Viral_helicase1; 1.		
DR	Pfam; PF03171; 2OG-FeII_Oxy; 1.		
KW	ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;		
KW	Transferase.		
FT	NP_BIND 915 922 ATP (POTENTIAL).		
SQ	SEQUENCE 1718 AA; 194529 MW; 820FEAE1E862415C CRC64;		
Alignment Scores:			
Pred. No.:	9.61	Length:	1718
Score:	73.50	Matches:	24
Percent Similarity:	44.30%	Conservative:	11
Best Local Similarity:	30.38%	Mismatches:	31

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan D., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarry C., McLeod M.P., McPherson D.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Renington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RL [5]
RL FUNCTION.
RX MEDLINE=97415323; PubMed=9271122;
RA Royzman I., Whittaker A.J., Orr-Weaver T.L.;
RT "Mutations in *Drosophila* DP and E2F distinguish G1-S progression from
RT an associated transcriptional program.";
RL Genes Dev. 11:1999-2011(1997).
RL [6]
RL FUNCTION.
RX MEDLINE=98078671; PubMed=9418862;
RA Duronio R.J., Bonnette P.C., O'Farrell P.H.;
RT "Mutations of the *Drosophila* dDP, dE2F, and cyclin E genes reveal
RT distinct roles for the E2F-DP transcription factor and cyclin E
RT during the G1-S transition.";
RL Mol. Cell. Biol. 18:141-151(1998).
RL [7]
RL TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=96087653; PubMed=8537434;
RA Hao X.F., Alphey L., Bandara L.R., Lam E.W., Glover D.,
RA La Thangue N.B.;
RT "Functional conservation of the cell cycle-regulating transcription
RT factor DRTF1/E2F and its pathway of control in *Drosophila*
RT melanogaster.";
RL J. Cell Sci. 108:2945-2954(1995).
CC -!- FUNCTION: Transcriptional activator that binds to E2F sites.
CC Required for wild-type growth in mitotic and polytene tissues,
CC contributes to the expression of replication genes at the G1-S
CC transition and Cyclin E.
CC -!- SUBUNIT: HETERODIMER OF E2F AND DP. COOPERATE TO GIVE SEQUENCE-
CC SPECIFIC DNA BINDING AND OPTIMAL TRANS-ACTIVATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Segmentally repeated expression throughout
CC early embryos is restricted to the ventral nerve cord in later
CC embryos.
CC -!- DEVELOPMENTAL STAGE: Throughout embryonic development.
CC -!- SIMILARITY: BELONGS TO THE E2F/DP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X78421; AAA55186.1; -;
DR EMBL; U10184; AAA19003.1; -;
DR EMBL; AB011813; BAA32746.1; -;
DR EMBL; AB003735; AAF55904.1; -;
DR HSP; Q16254; ICF7.
DR TRANSFAC; T01547; -;
DR FlyBase; FBgn0011766; E2f.
DR InterPro; IPR003316; E2F_TDP.
DR Pfam; PF02319; E2F_TDP; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Developmental protein; Activator.
FT DNA_BIND 253 318 POTENTIAL
FT DOMAIN 318 411 DIMERIZATION (POTENTIAL).
FT DOMAIN 318 411 POLY-SER.
FT DOMAIN 64 68 POLY-ASN.
FT DOMAIN 115 125 POLY-ALA.
FT DOMAIN 129 143 GLY-RICH.
FT DOMAIN 245 249 POLY-SER.
FT DOMAIN 519 573 GLY-RICH.
FT DOMAIN 525 533 POLY-GLN.
FT DOMAIN 594 601 POLY-ALA.
FT DOMAIN 701 710 POLY-GLY.
FT CONFLICT 127 127 H -> Q (IN REF. 1 AND 2).
SQ SEQUENCE 805 AA; 87460 MW; BD49C813DDB2A77D CRC64;

Alignment Scores:
Pred. No.: 11.2 Length: 805
Score: 72.50 Matches: 38
Percent Similarity: 37.29% Conservative: 28
Best Local Similarity: 21.47% Mismatches: 70
Query Match: 9.14% Indels: 41
DB: 1 Gaps: 5

US-09-856-221-1 (1-449) x E2F_DROME (1-805)
QY 34 TATTTTCATTGGATACGATTGTCAGAACCCGTTTAGAAGAGCGCCAGTCAAACC 93
DB 104 TTYLSErLeuAspLysGlyHisAsnThrGlyAlaValAlaThraAlaAlaAla 123
QY 94 ATTCGTGATCATTAATCAGCAATAAATCGGCAA-----CTGCGTCAA 135
DB 124 ThrAlaGlyHisThrGlnGlnGlnGlnHisHisGlnGlnGlnGln 143
QY 136 CGCGCGGCC-----TTGCCGCGGAACGTACTGCA 165
DB 144 ArgLysAlaThrGlyLysSerAsnAspIleThrAsnTyrTyrLysValLysArgPro 163
QY 166 ANTCTGTTAAACCGCTTTGTTCTTCCTCAGATAAACAACAAAGTACTGGCAG 225
DB 164 HisAlaValSerAspGluIleHisProLysLysGlnAlaLysGlnSerAlaHisGln 183
QY 226 ACCTTAGCACACGCG-----CTATAT 246
DB 184 ThrValTyrGlnLysHisThrAlaSerSerAlaProGlnGlnLeuArgHisSerHisHis 203
QY 247 AACTTACGTCAATACTGCAATTTGATGTCAGCCGCTGCTCATTCACCATCTATGGACA 306
DB 204 GlnLeuArgHisAspAlaAspAlaGluLeuAspGluAspValGluArgValAlaLys 223
QY 307 CCAGCAGAT-----CCGTCGCTACTGCTAGTGTGCGGTC 342
DB 224 ProAlaSerHisHisProPheSerLeuSerThrProGlnGlnLeuAlaAlaSerVal 243
QY 343 ACCGCTCACACGCGGAGGAGTTTCCTCGGACAGTAATGCCGATG----- 390
DB 244 AlaSerSerSerSerGlyAspArgAsnArgAlaAspThrSerLeuGlyIleLeuThr 263
QY 391 TACCGTTTTCGGATTATTCTCGAAATATGCCAAGTGGGGAGTACCCCACTG 441
DB 264 LysLysPheValLeuLeuGlnGlnSerProAspGlyValValAspLeu 280

Thu Jan 16 09:33:19 2003

CC -|- CATALYTIC ACTIVITY: Succinyl-CoA + dihydrolipoamide -> CoA + S-
 CC succinylhydrolipoamide.
 CC -|- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
 CC COFACTOR.
 CC -|- PATHWAY: Tricarboxylic acid cycle.
 CC -|- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
 CC SYMMETRY.
 CC -|- SUBCELLULAR LOCATION: Mitochondrial.
 CC -|- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
 CC -|- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D90401; BAA14397.1; -;
 DR PIR; A41015; A41015.
 DR HSP; P07016; 1CAT.
 DR InterPro: IPR001078; 2oxoacid_dh.
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR InterPro: IPR003016; Lipoyl.
 DR Pfam; PF00198; 2-oxoacid_dh; 1.
 DR Pfam; PF00364; biotin_lipoyl; 1.
 DR ProDom; PD001115; 2oxoacid_dh; 1.
 DR PROSITE; PS00189; LIPOYL; 1.
 KW Tricarboxylic acid cycle; Transferase; Acyltransferase; Lipoyl;
 KW Mitochondrion; Transit peptide.
 FT TRANSIT 1 56 MITOCHONDRION.
 FT CHAIN 57 442 DIHYDROLIPOAMIDE SUCCINYLTTRANSFERASE
 FT COMPONENT OF 2-OXOGLUTARATE DEHYDROGENASE
 FT COMPLEX.
 FT BINDING 99 99 LIPOYL (POTENTIAL).
 FT ACT_SITE 413 413 POTENTIAL.
 FT ACT_SITE 417 417 POTENTIAL.
 SQ SEQUENCE 442 AA; 47412 MW; 5109DB7F1DDFC2F6 CRC64;

Alignment Scores:
 Pred. No.: 11.8 Length: 442
 Score: 72.00 Matches: 27
 Percent Similarity: 40.40% Conservative: 13
 Best Local Similarity: 27.27% Mismatches: 37
 Query Match: 9.08% Indels: 22
 DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x ODO2_RAT (1-442)

QY	64	CCCCGTTT	AAGACCCG	CAGTCA	ACCATTCG	TGATCAT	TATATC	AGCATA	AAATGCGG	123
Db	198	ProLeu	alaglu	alaglu	alaglu	alaglu	alaglu	alaglu	alaglu	217
QY	124	CAACTGG	CGTCA	CGCGCC	TTGCGG	CGAAGC	TACTGC	AAATTC	GTTCGTT	183
Db	218	argMet	argGln	argIle	alaglu	argLeu	lysGlu	alaglu	asn-----	235
QY	184	TTCTTCT	CTCAG	TAAACA	AAAACT	GCAAAG	TACTGC	GACAGC	TTAGCAC	240
Db	236	MetLeu	thrThr	Phe	asnGlu	Val	aspMet	Ser	AsnIle	255
QY	241	-----	CTATAT	AACTAG	CTAAT	ATCTG	CAATG	TGTC	AGCCG	294
Db	256	Lys	aspAla	Phe	Leu	lysLys	His	AsnLeu	lysLeu	268
QY	295	ATCTAT	CGCACC	ACGAC	AGATCG	GTCTAG	CTGTAG	TCGTC	CGCTCA	351
Db	269	-----	-----	-----	-----	-----	-----	-----	-----	277

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: January 15, 2003, 15:44:13 ; Search time 42.2998 Seconds
(without alignments)
4374.264 Million cell updates/sec

Title: US-09-856-221-1

Perfect score: 793
Sequence: 1 tgcagcagcactcacttat.....gtgacccaactgatacagtt 449

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_plus.n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09856221/runat_15012003.153925.12457/app_query.fasta_1.1948
-DB=SPTREMBL_21 -OEW=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221.ecgn_1_171 -runat_15012003.153925.12457 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGESUBERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMPOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	86.3	2523	2	Q93RP0 xenorhabdus

	2	434.5	54.8	2516	2	Q9RN43
3	416.5	52.5	2376	2	Q9F9Z3	Q9RN43 photorhabdu
4	415	52.3	2504	2	Q85160	Q9F9Z3 serratia en
5	335	42.2	2538	2	Q93RN7	Q85160 xenorhabdu
6	227	28.6	1189	2	Q85152	Q93RN7 xenorhabdu
7	208	26.2	1565	2	Q85156	Q85152 photorhabdu
8	164	20.7	1279	2	Q93RP4	Q85156 xenorhabdu
9	84.5	10.7	270	16	Q8YZP6	Q93RP4 xenorhabdu
10	84.5	10.7	568	16	Q929G0	Q8YZP6 anabaena sp
11	84	10.8	338	16	Q8U8D8	Q929G0 chlamydia p
12	81	10.2	407	5	Q8T9A8	Q8U8D8 agrobacteri
13	81	10.2	578	5	Q8VMT7	Q8T9A8 drosophila
14	80.5	10.2	1275	17	Q8ZTL5	Q8VMT7 drosophila
15	79	10.0	523	16	Q8XS16	Q8ZTL5 pyrobaculum
16	78.5	9.9	682	2	Q9LAV1	Q8XS16 raistonia s
17	78	9.8	183	8	Q8WKU9	Q9LAV1 streptomyce
18	78	9.8	468	3	Q9URQ0	Q8WKU9 pardanthops
19	78	9.8	640	12	Q56294	Q9URQ0 candida alb
20	77.5	9.9	307	16	Q06723	Q56294 human herpe
21	77.5	9.8	1881	12	Q9DUN3	Q06723 bacillus su
22	77	9.9	509	17	Q8TWC3	Q9DUN3 vesicular e
23	77	9.7	535	5	Q26628	Q8TWC3 methanopyru
24	77	9.7	1944	5	Q9N6N2	Q26628 strongyloce
25	77	9.7	3201	5	Q9W0U2	Q9N6N2 leishmania
26	76.5	9.6	315	2	Q93SS3	Q9W0U2 drosophila
27	76.5	9.6	1240	12	Q9DWH8	Q93SS3 plestononas
28	76	9.6	457	5	Q9HH00	Q9DWH8 rat cytomeg
29	75.5	9.5	599	10	Q94E84	Q9HH00 leishmania
30	75.5	9.5	298	5	Q95P55	Q94E84 oryza sativ
31	75	9.5	260	11	Q91YM5	Q95P55 aedes aegypt
32	75	9.5	298	2	Q9RHLO	Q91YM5 mus musculu
33	75	9.5	466	3	Q9C1T1	Q9RHLO rhodopseudo
34	75	9.5	537	5	Q8SQS3	Q9C1T1 aspergillus
35	75	9.6	635	10	Q9SH38	Q8SQS3 encaphalito
36	75	9.5	1014	11	Q8VHJ7	Q9SH38 arabidopsis
37	74.5	9.4	407	16	Q97089	Q8VHJ7 mus musculu
38	74.5	9.4	452	2	Q59588	Q97089 streptococc
39	74.5	9.4	522	5	Q97773	Q59588 streptococc
40	74.5	9.4	670	5	Q9NGC4	Q97773 branchiosto
41	74.5	9.4	709	12	Q9Q0Q5	Q9NGC4 leishmania
42	74.5	9.4	715	12	Q9Q0Q4	Q9Q0Q5 influenza a
43	74.5	9.4	1437	5	Q44329	Q9Q0Q4 hiruoto medi
44	74	9.3	558	3	Q96T02	Q44329 pleurotus o
45	74	9.3	1186	5	Q17786	Q96T02 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q93RP0	PRELIMINARY;	PRT;	2523 AA.
AC	Q93RP0			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DE	Xptal protein.			
GN	XPTAL			
OS	Xenorhabdus nematophilus.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Xenorhabdus.			
OX	NCBI_TaxID=628;			
EN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PMF1296;			
RA	MEDLINE=21218513; PubMed=11319082;			
RT	Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;			
RT	"Sequence analysis of insecticidal genes from Xenorhabdus nematophilus			
RT	pmf1296.";			
RL	Appl. Environ. Microbiol. 67:2062-2069(2001).			
DR	EMBL; AJ308438; CAC38401.1; -			
DR	InterPro; IPR000953; Chromo.			
DR	SMART; SM00298; CHROMO; 1.			
SQ	SEQUENCE 2523 AA; 286999 MW; 3159852E0655B5B1 CRC64;			

```
Alignment Scores: 1.06e-68 Length: 2523
Pred. No.: 684.00 Matches: 134
Score: 684.00 Conservative: 6
Percent Similarity: 93.96% Mismatches: 9
Best Local Similarity: 89.93% Indels: 1
Query Match: 86.25% Gaps: 0
DB: 2

US-09-856-221-1 (1-449) x Q93RP0 (1-2523)
QY 2 GTGCAGCCTACCTTATTTGGCGATAA-CCTTATTTTTCATTTGGATAACGATTCGGTCA 60
DB 1903 ValGlnAlaLeuThrLeuLeuGlyAspGluProTyPheSerLeuAspAsnAspTrpSer 1922
QY 61 GAACCCCGTTTGAAGAAGCGCGCAGTCAAAACATTCGTGATCATTTATCAGCAATAAATG 120
DB 1923 GluProArgLeuGluGluAlaSerGlnThrMetArgHisHisTyrGlnHisLysMet 1942
QY 121 CGGCACTCGGTCAACGCGCGCTTCCGCGCAACCTACTGCAAAATTCGTTAACCGCT 180
DB 1943 LeuGlnLeuArgGlnArgAlaAlaLeuProThrLysArgThrAlaAsnSerLeuThrAla 1962
QY 181 TTGTTCCTTCCTCAGATAAACAACAACTGCAAAAGTTACTGGCAGAGCTTAGCAACAGC 240
DB 1963 LeuPheLeuProGlnIleAsnLysLysLeuGlnGlyTyrTrpGlnThrLeuThrGlnArg 1982
QY 241 CTATATACTTACGTACATAATCTGACAAATGTGATGTCAGCGCTTGTCTATACCCATCTAT 300
DB 1983 LeuTyrAsnLeuArgHisAsnLeuThrIleAspGlyGlnProLeuSerLeuLeuTyr 2002
QY 301 CGGACACACAGATCCCTCGCTACTGCTAGTCGTCGCGCTCAGCGCTCAGAGCGGA 360
DB 2003 AlaThrProAlaAspProSerMetLeuSerAlaIleThrAlaSerGlnGlyGly 2022
QY 361 GGGATTTCCTGGCAGATTAATCCGATACCGTTTTCGATTATTCGGAATATGCC 420
DB 2023 GlyAspLeuProHisAlaValMetProMetTyrArgPheProValIleLeuGluAsnAla 2042
QY 421 AAGTGGGAGTGACCAACTGATACAG 447
DB 2043 LysTrpGlyValSerGlnLeuIleGln 2051

RESULT 2
Q9RN43 PRELIMINARY; PRT: 2516 AA.
AC Q9RN43; 2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Toxin A (Toxin complex protein).
GN TCDA.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-W14;
RA Merlo D.J., Weigert L.M., Roberts J.L., Petell J.K.;
RT "Expression of Photorhabdus luminescens tcda gene in maize confers
corn rootworm tolerance.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-W14;
RX MEDLINE=21185117; PubMed=11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA French-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF188483; AAF05542.1; -.
DR EMBL; AF346500; AAL18486.1; -.

DR InterPro: IPR001589; Actbind_actnin.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
SQ SEQUENCE 2516 AA; 482952 MW; 8250A0650B614B99 CRC64;

Alignment Scores: 2.63e-40 Length: 2516
Pred. No.: 434.50 Matches: 87
Score: 434.50 Conservative: 20
Percent Similarity: 72.30% Mismatches: 38
Best Local Similarity: 58.78% Indels: 3
Query Match: 54.79% Gaps: 1
DB: 2

US-09-856-221-1 (1-449) x Q9RM43 (1-2516)
QY 13 CACCTTATTTGGCGGATAACCTTATTTTTCATTTGGATAACGATTTGGTACAGACCCCGTTTA 72
DB 1888 HisLeuLeuGlyAspLysProTyPheSerLeuProLeuSerThrThrTrpSerAspProArgLeu 1907
QY 73 GAAGAAGCGCGCAGTCAAACTTCGTGATCATTTATCAGCATATAAATGCGCAACTCGGT 132
DB 1908 AspArgAlaAlaAspIleThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927
QY 133 CACGCGCGCGCTTCCGCGGAAA-----CGTACTGCAAAATTCGTTAACCGCTTTG 183
DB 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947
QY 184 TTCCTTCCTCAGATAAACAACAACTGCAAAAGTTACTGGCAGAGCTTAGCAACAGCCCTA 243
DB 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967
QY 244 TATAACTTACGTACATAATCTGACAAATTTGATGTCAGCGCTTGTCTATACCATCTATCGC 303
DB 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987
QY 304 ACACGACGAGATCCGTCGCTACTGCTAGTCGCTGCTCAGCGCTCAGAGCGGAGG 363
DB 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaIleValAlaThrSerGlnGlyGly 2007
QY 364 GATTTCCCTCGGACAGTAATGCCGATGATACCGTTTTCGATTATTCGGAATATGCCAAG 423
DB 2008 LysLeuProGluSerPheMetSerLeuTyrArgPheProHisMetLeuGluAsnAlaArg 2027
QY 424 TGGGGAGTGACCAACTGATACAG 447
DB 2028 GlyMetValSerGlnLeuThrGln 2035

RESULT 3
Q9F9Z3 PRELIMINARY; PRT: 2376 AA.
AC Q9F9Z3; 2000 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE SepA.
DE SepA.
GN SepA.
OS Serratia entomophila.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=42906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ALMO2;
RX MEDLINE=20416224; PubMed=10960097;
RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
RT "plasmid-located pathogenicity determinants of Serratia entomophila,
the causal agent of amber disease of grass grub, show similarity to
the insecticidal toxins of photorhabdus luminescens.";
RL J. Bacteriol. 182:5127-5138(2000).
DR EMBL; AF135182; AAG09642.1; -.
KW Plasmid.
SQ SEQUENCE 2376 AA; 262649 MW; E418DACE22DBB7BF CRC64;
```

DR	InterPro:	IPR000566; LipocIn_CyLPABP.
SQ	PROSITE:	PS00213; LIPOCALIN; UNKNOWN_1.
SQ	SEQUENCE:	2504 AA; 280654 MW; 8*29B1693D1047CE CRC64;
Alignment Scores:		
Pred. No.:	4,35e-38	Length: 2504
Score:	415.00	Matches: 90
Percent Similarity:	71.33%	Conservative: 47
Best Local Similarity:	60.00%	Mismatches: 11
Query Match:	52.33%	Indels: 3
DB:	2	Gaps: 1
US-09-856-221-1 (1-449) x 085160 (1-2504)		
QY	5	CAGGCACCTACCTTTATTGGGCGCATAA-CCTTATTTTCATTCGATAACGATTGCTCAGAA 63
Db	1880	GlnAlaLeuAsnLeuGlyAspGluProGlnValMetLeuSerThrThrTrpAlaAsn 1899
QY	64	CCCCGTTTAGAAGAACGCCCGAGTCAAAACCATTCGTGATCATATACGCATAAATGCGG 123
Db	1900	ProThrLeuGlyAsnAlaLaSerIysThrThrGlnGlnValArgGlnValLeuThr 1919
QY	124	CAACTGCGT-----CAACGCGCGCCCTTCGCGCGAAACGTACTGCCAAATTCGTTAAC 177
Db	1920	GlnLeuArgLeuAsnSerArgValIysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1939
QY	178	GCTTTGTCCTTCAGATAAACAACAACTGCAAAAGTTACTGGCAGACGTTAGCAGAA 237
Db	1940	AlaLeuPheLeuProGlnGluAsnSerIysLeuIysGlyTrpTrpArgThrLeuAlaGln 1959
QY	238	CGCTATATACTTACGTTCATATCTGACAATTGATGTCAGCGCGTGTGCATTACCCATC 297
Db	1960	ArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuSerLeuProLeu 1979
QY	298	TATGGACACACGACGATCCGTCGTACTGCTTAGTGTCGCGTCACGCCCTCACAAGGC 357
Db	1980	TyrAlaIysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly 1999
QY	358	GGAGGGATTTGCCCTCGGACAGTAATCCGCATGTACCGGTTTTCGGATTATTCGGAAAT 417
Db	2000	GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGluGly 2019
QY	418	GCCAACTGGGGAGTGACCAACTGATACAG 447
Db	2020	AlaArgGlyLeuValAsnGlnLeuIleGln 2029
RESULT 5		
ID	Q93NR7	PRELIMINARY; PRT; 2538 AA.
AC	Q93NR7;	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	Xp2A2 protein.	
GN	Xp2A2.	
OS	Xenorhabdus nematophilus.	
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC	Xenorhabdus.	
OX	NCBI_TaxID=628;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=PMFI296;	
RX	MEDLINE=21218513; PubMed=11319082;	
RA	Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;	
RT	"Sequence analysis of insecticidal genes from Xenorhabdus nematophilus	
RT	PMFI296.";	
RL	Appl. Environ. Microbiol. 67:2062-2069(2001).	
DR	EMBL; AJ308438; CAC3804.1; -;	
SQ	SEQUENCE 2538 AA; 283995 MW; 3A7A98CF54FD756E CRC64;	
Alignment Scores:		
Pred. No.:	5.55e-29	Length: 2538
Score:	335.00	Matches: 76

Percent Similarity: 61.07% Conservative: 15
 Best Local Similarity: 51.01% Mismatches: 58
 Query Match: 42.24% Indels: 1
 DB: 2 Gaps: 0

US-09-856-221-1 (1-449) x Q93RW7 (1-2538)

QY 2 GTGAGGCACTCACCTTATTGGGGGATAA-CCTTATTTTCATTGGATAACGATTGGTCA 60
 DB 1918 ValArgAlaLeuGluLeuLeuGlyAspGluProGluAspTyrGlySerGlnGlnTrpAla 1937

QY 61 GAACCCGGTTTACAGAAGCCGCGAGCAAAACCACTGCGATCATATACAGATAAAAG 120
 DB 1938 AlaProSerLeuSerValAlaAlaSerHisThrValGlnAlaAlaTyrGlnGlnAspLeu 1957

QY 121 CGGCACTGGTCAACGGCGCGCTTCCGCGGAAACGACTGCAAAATTCGTTAAACCGCT 180
 DB 1958 ThrAlaLeuAspAsnGlyGluValAlaThrGlnProArgThrAlaAsnSerLeuValGly 1977

QY 181 TTGTTCCTTCCTCAGATAAACAACAACTGCAAGTTACTGGCAGAGCTTAGCACACGC 240
 DB 1978 LeuPheLeuProGluTyrAsnProAlaLeuThrAspTyrTrpGlnThrLeuAlaGLeuArg 1997

QY 241 CTATATACCTACCTCATAATCTCACAAATGATGTCACCGCTGTCTATACCCATCAT 300
 DB 1998 LeuPheAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuSerLeuAlaIleTyr 2017

QY 301 GCGACACGACAGATCCGTCGCTACTGCTTAGTGTGCGGTCAACGCGCTCACAAAGCGGA 360
 DB 2018 AlaGluProThrAspProlyAlaLeuLeuProSerMetValGlnAlaSerGlnGlyGly 2037

QY 361 GGGGATTGCTCGGACACTAATGCGCATGCTACCGTTTCCGATTATTCGAAATGCG 420
 DB 2038 SerAlaValLeuProGlyThrLeuSerLeuTyrArgPheProValMetLeuGluArgAla 2057

QY 421 AAGTGGGGATGACCACTCATACAG 447
 DB 2058 ArgAsnLeuValAlaGlnLeuThrGln 2066

RESULT 6
 ID 085152 PRELIMINARY; PRT; 1189 AA.
 AC 085152;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Insecticidal toxin complex protein TcCB.
 GN TCAB.
 OS Photobacterium luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Photobacterium.
 OX NCBI_TaxID=29488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-W-14;
 RX MEDLINE=98306208; PubMed=9641921;
 RA Bowen D., Rochelleau T.A., Blackburn M., Andreev O., Golubeva E.,
 RA Bhartiya R., French-Constant R.H.;
 RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";
 RL Science 280:2129-2132(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-W14;
 RX MEDLINE=21185117; PubMed=11286884;
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
 RA French-Constant R.H.;
 RT "The tc genes of Photobacterium: a growing family.";
 RL Trends Microbiol. 9:185-191(2001).
 DR EMBL; AF046867; AAC38624.1; -;
 DR EMBL; AF348497; AAL18450.1; -;
 DR InterPro; IPR002927; Virus_HS.
 DR Pfam; PF01550; Virus_HS; 1.
 SQ SEQUENCE 1189 AA; 131466 MW; 7DAA3AE0AB88127 CRC64;

Alignment Scores:
 Pred. No.: 8,92e-17 Length: 1189
 Score: 227.00 Matches: 57
 Percent Similarity: 49.01% Conservative: 17
 Best Local Similarity: 37.75% Mismatches: 65
 Query Match: 28.63% Indels: 12
 DB: 2 Gaps: 3

US-09-856-221-1 (1-449) x 085152 (1-1189)

QY 16 CTTATTGGCGGATACCTTATTTTCATTGGATAACGATTGGTCAACCCCTTTAGAA 75
 DB 583 LeuLeuGlyProArgProAspIleHisThrAsnThrTrpProAsnProThrLeuSer 602

QY 76 GAAGCGCGCAGTCAAAACCATTCGTGATCATATACGATATAAATGCGGCACTGCGTCAA 135
 DB 603 LysGluAlaGlyAlaIleAlaThrProThrPheLeuSerSerProGluValMetThrPhe 622

QY 136 CGCGCGCGCTTCCGCGGAAACGTTACTGCAAAATTCGTTAACCGCTTTGTTCCTCTCAG 195
 DB 623 AlaAlaTrpLeuSerAlaGlyAspThrAlaAsnIleGlyAspGlyAspPheLeuProPro 642

QY 196 ATAAACAAAAAAGTACTGCGCAGAGCTTAGCACAAACCCCTATATAACTTACGT 255
 DB 643 TyrAsnAspValLeuLeuGlyTyrTrpAspLysLeuGluLeuArgLeuTyrAsnLeuArg 662

QY 256 CATATCTGACAAATGATGTCAGCGCTTGTCTATACCATCTATGCGACACGACAGAT 315
 DB 663 HisAsnLeuSerLeuAspGlyGlnProLeuAsnLeuProLeuTyrAlaThrProValAsp 682

QY 316 CCGTCGCTACTGCTTAGTGTGCTGCGCTCACAGCGCTCACAAAGCGGAGGGAT----- 366
 DB 683 ProLysThrLeu-----GlnArgGlnGlnAlaGlyGlyAspGlyThrGly 697

QY 367 ---TTCCT-----CGACAGTAATGCGGATGTACCGTTTCCGATTATTCGGAA 414
 DB 698 SerSerProAlaGlyGlyGlnGlySerValGlnGlyTyrArgTyrProLeuLeuValGlu 717

QY 415 AATGCCAAGTGGGAGTGACCCCACTGATACAG 447
 DB 718 ArgAlaArgSerAlaValSerLeuLeuThrGln 728

RESULT 7
 ID 085156 PRELIMINARY; PRT; 1565 AA.
 AC 085156;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Insecticidal toxin complex protein TcCB.
 GN TCBB.
 OS Photobacterium luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Photobacterium.
 OX NCBI_TaxID=29488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-W-14;
 RX MEDLINE=98306208; PubMed=9641921;
 RA Bowen D., Rochelleau T.A., Blackburn M., Andreev O., Golubeva E.,
 RA Bhartiya R., French-Constant R.H.;
 RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";
 RL Science 280:2129-2132(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-W14;
 RX MEDLINE=21185117; PubMed=11286884;
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
 RA French-Constant R.H.;
 RT "The tc genes of Photobacterium: a growing family.";
 RL Trends Microbiol. 9:185-191(2001).
 DR EMBL; AF046867; AAC38624.1; -;
 DR EMBL; AF348497; AAL18450.1; -;
 DR InterPro; IPR002927; Virus_HS.
 DR Pfam; PF01550; Virus_HS; 1.
 SQ SEQUENCE 1189 AA; 131466 MW; 7DAA3AE0AB88127 CRC64;


```

Alignment Scores:
Pred. No.: 1.01 Length: 270
Score: 84.50 Matches: 40
Percent Similarity: 40.96% Conservative: 14
Best Local Similarity: 34.78% Mismatches: 34
Query Match: 10.66% Indels: 27
DB: 16 Gaps: 8

US-09-856-221-1 (1-449) x Q8YZP6 (1-270)
QY 91 ACCATTGCTGATCATATTCATGCAATGCGCAACGTCGCAACGGCGCTTTCGCG 150
Db 3 ThrileserAspHisphgln---SerLeuArgGlnArgGlnGlnCysAlaLeuLeuPro 21
QY 151 GCGAAGCTACTCGAATTCGTTACCGCTTTCTTCCTCAGATAAACAACAACTG 210
Db 22 -----PheileThrAlaGlyAspProAspLeu 30
QY 211 CAAAGTTACTGGCAGAGCTTAGCACAACGCTATATAACTTA---CGTCATAATCTGACA 267
Db 31 -----GlnThrThrAlaGluAlaLeuArgIleLeuAspArgHisGly----- 44
QY 268 ATTGATGTCAGCGCTTCATTACCTATCTATGCGACACCA---GCAGATCCGTCCTGA 324
Db 45 AlaaspPheileGluLeuGlyValPro---TyrSerAspProLeuAlaAspGlyProVal 63
QY 325 CTGCTTAGTCTGCTACCGCTCACCCTCACAAGCGCGAGGGATTCGCCCTCGG----- 375
Db 64 ileGlnAlaAlaIaThrArgAlaLeuGlnArgGlyThrLysLeuGluGlnValLeuGlu 83
QY 376 -----ACAGTAATGCCGATGTACCGTTTCCGATTATTCTG 411
Db 84 MetLeuGluThrValIleProSerLeuGlnAlaProIleLeu 98

RESULT 10
QY29G0 PRELIMINARY; PRT; 568 AA.
AC QY29G0:
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE Putative OMP protein.
GN CPN0021 OR CPJ0021 OR CP0755.
OS Chlamydia pneumoniae (Chlamydiales: Chlamydiaceae: Chlamydiales).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192398;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.S.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kubara S., Nakazawa T.;

```

Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.:

BL Nucleic Acids Res. 28:2311-2314(2000).

DR EMBL; AF001587; AAD18174.1; -.

DR EMBL; AF002234; AAF38558.1; -.

DR EMBL; AF002545; BAA98233.1; -.

DR TIGR; CP0755; -.

KW Complete proteome.

SQ SEQUENCE 568 AA; 63553 MW; 3358FCB5C482E80 CRC64;

Alignment Scores:

Pred. No.: 1.22 Length: 568

Score: 84.50 Matches: 40

Percent Similarity: 42.66% Conservative: 21

Best Local Similarity: 27.97% Mismatches: 55

Query Match: 10.66% Indels: 27

DB: 16 Gaps: 6

US-09-856-221-1 (1-449) x Q929G0 (1-568)

QY 58 TCAGAACCCCGCTTTAGAGAAGCC-----GCCAGTCAAAACCATTCGTGAT----- 102

Db 232 AlaSerProGlnAspGlnGluAlaIleLeuTyrAlaLeuGlyLysLeuLysAspGlyGln 251

QY 103 CATTATCAGCATAAATCGCGCAACTGCGTCAACCGCGCGCTTCCGCGGAACGTACT 162

Db 252 SerTyrTyrAsnIleLysLysGlnLeuGlnLysProAspValAspValThrLeuAla 271

QY 163 GCAATTCGTTAACCGCTTTG-----TTCTCTCTCAGATAAACAACAAA 207

Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluGluAspAlaLeuProValIleLysLys 290

QY 208 CTGCAAAAGTTACTGCGCAGAGTTAGCACACGC-----CTATATAACTTACGTCTAT 258

Db 291 -----GlnAlaLeuGluGluArgProArgAlaLeuTyrAlaLeuArgHis 305

QY 259 AATCTGACAATTCGTCAGCGCTTGCATTACCATCTATCGACACACCATCGG 318

Db 306 LeuProSerGluIleGlyIleProIleAlaLeuProIleLeuLysThrLysAsnSer 325

QY 319 TCGTACTGCTAGTGTGCTGCTCAGCGCTCAGACGGCGAGGGATTCGCT----- 372

Db 326 GluAlaLysLeuAsnValAlaLeuAlaLeuGluGlyLysAspThrProLysLeu 345

QY 373 -----CGCACAGTAATCGCGATGACCGGTTTCCGATTATTCTGGAA 414

Db 346 LeuGluTyrIleThrGluArgLeuValGlnProHisTyrAsnGluThrLeuAlaLeuSer 365

QY 415 AATGCCAAG 423

Db 366 PheSerLys 368

RESULT 11

QY29G0 PRELIMINARY; PRT; 338 AA.

ID QY29G0

AC QY29G0:

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DE ABC transporter, substrate binding protein.

GN ATU4154 OR AGR_L1399.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OX NCBI_TaxID=176299;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,

RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,

RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,

RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,

RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,

RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
RP "the genome of the natural genetic engineer Agrobacterium tumefaciens
C58";
RT C58.;
RL Science 294:2317-2323(2001).
[2]
RN
RP
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
Raoult B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
Quemell K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
Cielo C., Slater S.;
RP "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58";
RT Science 294:2323-2328(2001).
RL EMBL; AF009345; AAL44954.1; ALT_INIT.
DR EMBL; AF008268; AAK89280.1; .
KW Complete proteome.
SQ SEQUENCE 338 AA; 35571 MW; 879AA0BEFDD3C34F CRC64;

Alignment Scores:
Pred. No.: 1.22 Length: 338
Score: 84.00 Matches: 42
Percent Similarity: 43.70% Conservative: 17
Best Local Similarity: 31.11% Mismatches: 53
Query Match: 10.76% Indels: 23
DB: 16 Gaps: 8

US-09-856-221-1 (1-449) x Q8U88 (1-338)

QY 443 ATCACTGGTGGTCACT-----CCCCACTGGCATTTTCAGATTAATCGGAAA 396
DB 83 ValSerTrpValGluPheThrAlaGlyProLeuValGluAlaLeuAsnValGlySer 102
QY 395 CGGTACATCGGCATTACTCGCGAGGCAATCCCTCCG----- 357
DB 103 IleAsnValGlyTrpThr-----GlyAspAlaProIlePheGlyGlnAlaLagly 120
QY 356 CTTGTGAGCGGTGAGCGGCACCTAAGCAGTACGACGAGTCTCGTGGTCCCATAG 297
DB 121 SerAlaIleValTyrrValAlaAlaLeuProSer---AsnGlyLysGlyGluAlaIlePhe 139
QY 296 ATGGTAATGACACGCGTACCATCAATTCAGATTATGACGTAAGTATATAGCGGT 237
DB 140 ThrLysProGluSerGlyIleLysSerVal-AlaAspLeuLysGlyLysValGlyVa 159
QY 236 TGTGCTA---ACGCTGCCAGTAACCTTTCAGTTTGTATCTGAGGAAGGAACAA 180
DB 159 IclLysGlyLysSerAlaHisAsnLeuValAlaAlaLeuGluLysAsnGlyLeuLy 179
QY 179 GCGGTTAACCAATTTGCGAG-----TACGTTTCGCGGCAAGCGCGGTTGACGCGAGT 126
DB 179 s---LeuSerAspIleAspValThrTyrrLeuSerProAla-----AspAlaAla 194
QY 125 TCCGCCATTTATGCTCATTAATGATCAGCAAGTGGTTTGACTGG 83
DB 194 aAlaAlaPheAlaSerAspLysIleAspAlaTrpAlaValTrp 208

RESULT 12
Q8T9A8
ID Q8T9A8 PRELIMINARY; PRT; 407 AA.
AC Q8T9A8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SD10272p.
GN MSP-300.
OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069852; AAL3997.1; . 9E76C03821C0CD40 CRC64;
SQ SEQUENCE 407 AA; 47289 MW; 9E76C03821C0CD40 CRC64;

Alignment Scores:
Pred. No.: 2.8 Length: 407
Score: 81.00 Matches: 27
Percent Similarity: 41.28% Conservative: 18
Best Local Similarity: 24.77% Mismatches: 40
Query Match: 10.21% Indels: 24
DB: 5 Gaps: 3

US-09-856-221-1 (1-449) x Q8T9A8 (1-407)

QY 61 GAACCCCGTTTAGAGAA----- 78
DB 294 AspProArgLeuSerGluAsnValAspAspLeuValArgValLeuAspGlyLeuGlyGlu 313
QY 79 -----GCCGCGCAGTCAACCATTCGT-----GATCATTATCAG 111
DB 314 LeuValLysAlaLysSerGlnSerLeuGluGlnThrLeuAlaGlnIleAspValTyrrGln 333
QY 112 CATAAATGCGGCAACTCGCTCAACGCGCGGCTTCCCGCGCAACAGTACTGCAAAATTCG 171
DB 334 GlnGlnMetGlnSerLeuArgGlnArgIleLeuGlnGluGlnLeuArgLeuVal 353
QY 172 TTAACCGCTTTGTTCTCTCCCTCAGATAAACAACAACTGCAAGTACTGCGCAGCGTTA 231
DB 354 MetAlaProThrTyrrLeuProHisAspArgGluArgAlaLeuAlaGluGlnGlnAlaCys 373
QY 232 GCACACGCGTATATACTTACGTATATACTGACAAATTCGACAAATTCGTCACCGCTTCATTA 291
DB 374 ArgGluArgValLysAsnLeuHisSerLysIleThrAlaArgAsnGluArgIleLysLeu 393
QY 292 CCCATCTAT---GCGACACACGACAGAT 315
DB 394 LeuIleHisArgGlyThrProAspAsp 402

RESULT 13
Q9VMT7
ID Q9VMT7 PRELIMINARY; PRT; 578 AA.
AC Q9VMT7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG18251 protein.
GN MSP-300 OR CG18251 OR CG18252.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbavani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harlis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Welstock G.M., Welssbach J.,
 RA Williams S.M., Woodruff T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003608; AAF52224.1; -;
 DR FlyBase: FBgn010070; Msp-300.
 DR InterPro: IPR002017; Spectrin.
 DR SMART: SM00150; SPEC; 1.
 SQ SEQUENCE 578 AA; 67145 MW; 27B0BAD7224A685D CRC64;

Alignment Scores:
 Pred. No.: 3.06 Length: 578
 Score: 81.00 Matches: 27
 Percent Similarity: 41.28% Conservative: 18
 Best Local Similarity: 24.77% Mismatches: 40
 Query Match: 10.21% Indels: 24
 DB: 5 Gaps: 3

US-09-856-221-1 (1-449) x Q9VW7 (1-578)
 QY 61 GAACCCGTTTAGAGAA----- 78
 Db 465 AspProArgLeuSerGluAsnValAspLeuValArgValLeuAspGlyLeuGlu 484
 QY 79 -----GCCGCCAGTCAACACCATTCGT-----GATCATTATCAG 111
 Db 485 LeuValLysAlaLysSerGlnSerLeuGluInThrLeuAlaGlnIleAspValTyrGln 504
 QY 112 CATAAATCGGCAACTCGCTCAACGGCGGCTTCGCCGGAACGCTACTGCAAAATTCG 171
 Db 505 GlnGlnMetGlnSerLeuArgGlnIleGlnGlnGluGlnGlnLeuArgLeuVal 524
 QY 172 TTACCGCTTGTCTCTCAGATAAACAACAACTGCAAAAGTTACTGGCAGAGTTA 231
 Db 525 MetAlaProThrTyrLeuProHisAspArgGluArgAlaLeuAlaGluGlnAlaCys 544
 QY 232 GCACACGGCTATATACCTACGTCATAATCTGACAAATTCATGTCACGGTTGTCTATTA 291
 Db 545 ArgGluArgValLysAsnLeuHisSerLysIleThrAlaArgAsnGluArgIleLysLeu 564
 QY 292 CCATCTAT---GCCACACCGACAGAT 315
 Db 565 LeuIleHisArgGlyThrProAspAsp 573
 RESULT 14

Q8ZTL5
 ID Q8ZTL5 PRELIMINARY; PRT; 1275 AA.
 AC Q8ZTL5;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE ATP-dependent helicase, probable.
 GN PAE3198.
 OS Pyrobaculum aerophilum.
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
 OC Thermoproteaceae; Pyrobaculum.
 OX NCBI_TaxID=13773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
 RX PubMed=11792869;
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
 RA Miller J.H.;
 RT "Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
 DR EMBL: AF005916; AAL64745.1; -;
 DR InterPro: IPR001410; DEAD
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KW Helicase; Complete proteome.
 SQ SEQUENCE 1275 AA; 143418 MW; 866C3439310EC090 CRC64;

Alignment Scores:
 Pred. No.: 4.27 Length: 1275
 Score: 80.50 Matches: 25
 Percent Similarity: 44.44% Conservative: 19
 Best Local Similarity: 25.25% Mismatches: 36
 Query Match: 10.15% Indels: 19
 DB: 17 Gaps: 2

US-09-856-221-1 (1-449) x Q8ZTL5 (1-1275)
 QY 73 GAAGAAGCGCGCAACCAATTCGTGATCATATCAGCATAAATGCGCAACTGCGT 132
 Db 570 AspAspAlaLeuSerSerSerIleValThrAsnProTyrLysLeuSerMetLeu 589
 QY 133 CAACGCGCGCTTCGCGGGAACGCTACTGCAAAATTCGTTAACCCGCTTGTCTCT 192
 Db 590 MetArgSerAlaAspSerAlaLysLysLeuAlaAsnSerLeuAlaLeuValIleAsp 609
 QY 193 CAGATAAACAACAACTGCAA-----AGTTACTGGCAGAGCTTAGCAACAGC 240
 Db 610 GluValHisThrMetLeuGluProLysHisLeuAspPhePheThrAlaLeuHisArg 629
 QY 241 CTATATACTTACGTCATAATCTGACAATTGATGTCAGCGTTGTGTCATTACCATCAT 300
 Db 630 LeuTyrLeuLeuGlyAsp----- 635
 QY 301 GCACACGACGACGATCGCTCGTACTGCTTACTGCTGCTACCGCTCACAAGGC 357
 Db 636 ---ValLysLysTyrProAlaIleLeuSerSerAlaThrValThrSerGly 653
 RESULT 15
 Q8XS16
 ID Q8XS16 PRELIMINARY; PRT; 323 AA.
 AC Q8XS16;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Probable magnesium and cobalt transport transmembrane protein.
 GN RSP0865 OR RS01784.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salancubet M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandelier M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspier C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Squier P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646080; CAD17816.1; .
 DR InterPro; IPR002523; Cora.
 DR Pfam; PF01544; Cora.1
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 323 AA; 36251 MW; DE264559E199DD61 CRC64;

 Alignment Scores:
 Pred. No.: 4.45 Length: 323
 Score: 79.00 Matches: 46
 Percent Similarity: 37.57% Conservative: 22
 Best Local Similarity: 25.41% Mismatches: 66
 Query Match: 9.96% Indels: 47
 DB: 16 Gaps: 7

US-09-856-221-1 (1-449) x Q8XS16 (1-323)
 QY 3 TGCAGGCACTCACCTTAT-----TGGGCGATAACC 32
 DB 126 CysGluHisValProHisLeuAlaLysGlyAlaProPheAlaLeuTyrAlaValMet 145
 QY 33 TTATTTTTCATGATACGATTGGTCAGAACCCGCTTTAGAA-----GAAGCCGCCAG 86
 DB 146 AspPheValVal-AspAsn-----TyrGlnProValLeuGluSerMetGluAlaAlaPh 163
 QY 87 TCAACCATTCGTGATCAT----- 105
 DB 163 eAspAlaIleGluGlyGlnLeuPheGlyAspAlaPheAspArgAlaAlaIleGluArgLe 183
 QY 106 -TATCAGCATAAATCGCGCACTCGCTCAACGC---GCGGCTTGGCGGAAACGTAC 161
 DB 183 uTyrThrLeuLysArgGlnLeuLeuArgLeuArgAsnAlaAlaLeuProValGluAspIle 203
 QY 162 TGCAAATTCGTTAACCGCTTTGTTCTCTCAGATAAACAACAAAGTTCCTG 221
 DB 203 eAlaGlyGlnLeuValArgLeuHisGluAspValProLysGlnLeuArgAlaTyrPh 223
 QY 222 GCAGACGTTAGCACACGCCCTATATACTTACGTCATATCTGACAAATTGATGTCAGCC 281
 DB 223 eArgAspValAlaAspHisAlaHisArgLeuValGlyAlaLeuAspValIleArgGluMe 243
 QY 282 GTTGTCTATTACCATCTATCGACACGACGATCCGCTGCTTGTAGTGTGTCGCGT 341
 DB 243 tLeuThrAlaIleSerValAsnValAlaLeuValSerValThrGlnAsnAspIleVal 263
 QY 342 CACCGCCTCACAGGCGGAGG-----GATTTCCTCGGACAGTAAATGCCGATGTA 392
 DB 263 lLysArgLeuAlaGlyTrpGlyAlaIleLeuAlaIleProThrValValPheSerAsnTy 283
 QY 393 C-----CGTTTTCGCGATTATTCT 410
 DB 283 rGlyMetAsnPheLysGlyMetProGluLeuGluHisProAlaGlyTyrProIleValLe 303
 QY 411 G 411
 DB 303 u 303

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 15, 2003, 15:39:33 ; Search time 14.2569 Seconds
(without alignments)
2878.683 Million cell updates/sec

Title: US-09-856-221-2
Perfect score: 265
Sequence: 1 caaacattcgtatcattca.....gcagacgttagcacacgcc 154

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09856221/runat_15012003.153924.12438/app_query.fasta_1.1948
-DB=A-Genesec 101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=diuom62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221.ecgn_1_1_60_erunat_15012003.153924.12438 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A-Genesec 101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	137	51.7	2376	22	AA97694	SepA protein encod
2	126.5	47.7	2504	18	AAW17871	Phototransduc lum
3	126.5	47.7	2504	19	AAW56557	Toxin TcBa, encod
4	126.5	47.7	2504	22	AAW72610	Phototransduc t
5	126.5	47.7	2505	22	AAW72612	Modified Photot
6	118.5	44.7	2516	18	AAW17899	Phototransduc lum
7	118.5	44.7	2516	19	AAW56572	Toxin TcDa, encod
8	118.5	44.7	2516	22	AAW72609	Phototransduc t
9	118.5	44.7	2517	22	AAW72611	Modified Photot
10	118.5	44.7	2537	22	AAW72614	TcDa toxin-zein
11	115.5	43.6	2522	20	AAW33729	Phototransduc lum
12	113	42.6	573	18	AAW18303	Phototransduc lum
13	113	42.6	573	19	AAW56559	Toxin TcBaIII, enc
14	108	40.8	579	18	AAW18301	Phototransduc lum
15	108	40.8	579	19	AAW56574	Toxin TcDaIII, enc
16	63.5	24.0	886	22	ABB71328	Drosophila melanog
17	63	23.8	364	23	AAU99816	Escherichia coli M
C 18	61	22.8	114	22	AAU56070	Propionibacterium
19	60.5	22.8	658	22	ABB61728	Drosophila melanog
20	59.5	22.5	54	22	AAU44965	Propionibacterium
21	59.5	22.5	66	22	AAU39885	Propionibacterium
C 22	59	22.1	118	22	AAO00632	Human polypeptide
C 23	58.5	21.9	639	22	ABB5465	Drosophila melanog
C 24	58.5	21.9	676	22	ABB59964	Drosophila melanog
C 25	58.5	21.9	676	22	ABB66818	Drosophila melanog
C 26	57	21.3	125	21	AAAB42930	Human ORFX ORP294
C 27	57	21.3	254	21	AAAB24446	Human secreted pro
C 28	57	21.3	276	21	AAAB24947	Human ORFX ORP2711
C 29	57	21.3	299	21	AAAB23042	Human myeloid upre
C 30	57	21.3	316	22	AAU27957	Human contig polyp
C 31	57	21.3	318	22	ABG22136	Novel human diag
C 32	57	21.3	322	21	AAAB15549	Human immune syste
C 33	57	21.3	322	21	AAAB19729	Human SECX Clone 4
C 34	57	21.3	322	21	AAAB19729	Human PRO110 (UNQ
C 35	57	21.3	322	22	AAU27785	Human full-length
C 36	57	21.3	322	22	AAU29109	Human PRO polypept
C 37	57	21.3	322	22	AAU93665	Human polypeptide,
C 38	57	21.3	322	22	AAU93665	Protein of the inv
C 39	56.5	21.2	584	22	AAU96296	Putative P. abyssi
C 40	56.5	21.2	951	16	AAU75704	Eph-related CEK6.
41	56	21.1	355	22	AAU34438	E. coli cellular p
42	56	21.1	355	22	AAU34438	Escherichia coli p
43	56	21.1	446	21	AAU57299	Arabidopsis thalia
44	56	21.1	446	21	AAU61050	Arabidopsis thalia
45	56	21.1	483	21	AAU57298	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA97694
ID AA97694 standard; Protein; 2376 AA.
XX
AC AA97694;
XX
DT 19-JUN-2001 (first entry)
XX
DE DE
XX
XX SepA protein encoded by Serratia insecticidal protein complex gene.
KW Insecticidal protein complex; amber disease; insect; Coleoptera;
KW pesticide; SepA protein.
XX
XX Serratia sp.
XX
XX WO200116305-A2.
XX
XX PD 08-MAR-2001.
XX

```

PF 04-SEP-2000; 2000WO-NZ00174.
XX
XX
PR 02-SEP-1999; 99NZ-0337610.
XX
XX (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX
XX Glare TR, Hurst MRH, Jackson TA;
PI
XX WPI: 2001-169009/17.
DR
DR N-PSDB; AAA91292.
XX
XX New nucleic acid encoding a polypeptide useful as a pesticide
PT especially for Coleoptera -
PT
XX
XX Claim 24; Page 92-99; 109pp; English.
XX
XX This sequence represents the SepA protein encoded by the Serratia
CC insecticidal protein complex gene of the invention. The invention relates
CC to a gene encoding an insecticidal protein complex or a functional
CC fragment, a neutral mutation, or a homologue of the complex. The
CC polypeptides and nucleotides of the invention are used to induce amber
CC diseases or like conditions in insects, especially to those from the
CC order comprising Coleoptera, useful as a pesticide.
XX
XX Sequence 2376 AA;
XX
XX Alignment Scores:
XX Pred. No.: 7,23e-11 Length: 2376
XX Score: 137.00 Matches: 28
XX Percent Similarity: 66.67% Conservative: 6
XX Best Local Similarity: 54.90% Mismatches: 17
XX Query Match: 51.70% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-856-221-2 (1-154) x AAY97694 (1-2376)
XX
XX QY 1 CAACACATTCGTGATCATTTATCAGCATAAATCGCGCACTGCGTCACGCGCGCCTTG 60
XX
XX Db 1785 GluValThrArgArgAspTyrGlnGluAlaLeuAlaValArgGluValProAla 1804
XX
XX QY 61 CCGCGCAAGACTACTGCAATTCGTTACCGCTTTGTTCTTCCTCAGATAACAAAAA 120
XX
XX Db 1805 ProGluThrArgThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnGluVal 1824
XX
XX QY 121 CTGCAAGTTACTGGCAGACGTAGCACACGC 153
XX
XX Db 1825 LeuLysGlyTyrTrpGlnThrLeuAlaGlnArg 1835
XX
XX RESULT 2
XX AAW17871
XX ID AAW17871 standard; Protein; 2504 AA.
XX
XX AC AAW17871;
XX
XX XX 29-JAN-1998 (first entry)
XX
XX DE Photorhabdus luminescens insect toxin protein TcBa.
XX
XX KW Insecticide; insect; toxin; pest control; biological control;
XX Photorhabdus luminescens; TcBa; Southern corn rootworm;
XX Colorado potato beetle; Western corn rootworm; meal worm;
XX boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
XX cabbage looper; codling moth; corn earworm; European corn borer;
XX tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
XX Diptera, Dictyoptera; Acarina; Homoptera.
XX
XX OS Photorhabdus luminescens strain W-14 (ATCC 55397).
XX
XX PN W09717432-A1.
XX
XX XX 15-MAY-1997.
XX
XX PD
XX
XX PF 06-NOV-1996; 96WO-US18003.
XX
XX
XX 04-SEP-1996; 96US-0705484.
XX
XX 06-NOV-1995; 95US-0007255.
XX
XX 28-FEB-1996; 96US-0608423.
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
XX Petell J, Roberts JL, Rocheleau TA, Schoonover S;
XX Strickland JA;
XX
XX WPI: 1997-281022/25.
DR
DR N-PSDB; AAT68836.
XX
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
XX can be genetically engineered into insect larvae food and plants for
XX insect control
XX
XX Claim 8; Page 119-129; 276pp; English.
XX
XX This polypeptide comprises a specifically claimed 280.6 kDa
CC insecticidal toxin, TcBa, of Photorhabdus luminescens that is
CC proteolytically cleaved into 207.6 kDa (see AAW18302) and 62.9 kDa
CC polypeptides (see AAW18303). The TcBa polypeptide can be expressed
CC in host cells using a gene (see AAT68836) isolated from a genomic
CC library. Claimed toxin proteins of P. luminescens (see AAW17871,
CC AAW17884-89, AAW17899-900, AAW18301-06) can be applied to, or
CC genetically engineered into, insect larvae food and plants for insect
CC control. The Photorhabdus toxins are particularly effective against
CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,
CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black
CC cutworm, cabbage looper, codling moth, corn earworm, European corn
CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and
CC are also active against insects of the orders Hymenoptera, Diptera
CC Dictyoptera, Acarina and Homoptera. (All claimed).
XX
XX Sequence 2504 AA;
XX
XX Alignment Scores:
XX Pred. No.: 2,5e-09 Length: 2504
XX Score: 126.50 Matches: 28
XX Percent Similarity: 64.71% Conservative: 5
XX Best Local Similarity: 54.90% Mismatches: 17
XX Query Match: 47.74% Indels: 1
XX DB: 18 Gaps: 1
XX
XX US-09-856-221-2 (1-154) x AAW17871 (1-2504)
XX
XX QY 1 CAACACATTCGTGATCATTTATCAGCATAAATCGCGCACTGCGTCACGCGCGCCTTG 60
XX
XX Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929
XX
XX QY 61 CCGCGCAAGACTACTGCAATTCGTTACCGCTTTGTTCTTCCTCAGATAACAAAAA 120
XX
XX Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnSerLys 1949
XX
XX QY 121 CTGCAAGTTACTGGCAGACGTAGCACACGC 153
XX
XX Db 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960
XX
XX RESULT 3
XX AAW56557
XX ID AAW56557 standard; Protein; 2504 AA.
XX
XX AC AAW56557;
XX
XX XX 07-AUG-1998 (first entry)
XX
XX DT Toxin TcBa, encoded by the tcbA gene from genomic region tcb.
XX
XX DE Photorhabdus luminescens W-14; nematode; symbiotic;
XX Heterorhabditis; tcb; tcb; tcc; tcd; insecticidal activity; toxin;
XX

```


KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
KW Homoptera; Southern; western corn rootworm; Colorado potato beetle;
KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW Tobacco hornworm; budworm.
XX
OS Photorhabdus luminescens.
XX
PN WO9808932-A1.
XX
PD 05-MAR-1998.
XX
PF 05-MAY-1997; 97WO-US07657.
XX
PR 06-NOV-1996; 96WO-US18003.
PR 28-AUG-1986; 96US-0705484.
PR 06-NOV-1996; 96US-0743699.
XX
PA (DOWC) DOWELANCO.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI Strickland JA, Sukhapinda K;
XX
XX WPI; 1998-179427/16.
DR N-PSDB; AAV29985.
XX
XX Isolated toxins from Photorhabdus luminescens strains - useful for
control of insect pests
XX
PS Claim 34; Pages 163-169; 321pp; English.
XX
CC The present sequence represents a protein named TcBa of the bacterium
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
CC nematodes of the heterorhabditis genus. The bacterium has at least 4
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
CC produced from these regions that are associated with insecticidal
CC activity. The native toxins are secreted proteins. The proteins are
CC toxic to insects upon exposure and especially when ingested. The
CC nucleic acid sequence can be used to produce transgenic plants,
CC baculoviruses or microbial hosts for toxin production. They can be used
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC codling moth, corn earworm, European corn borer or tobacco hornworm
CC or budworm.
XX
SQ Sequence 2504 AA;
Alignment Scores:
Pred. No.: 2,5e-09 Length: 2504
Score: 126.50 Matches: 28
Percent Similarity: 64.71% Conservative: 5
Best Local Similarity: 54.90% Mismatches: 17
Query Match: 47.74% Indels: 1
DB: 19 Gaps: 1
US-09-856-221-2 (1-154) x AAV56557 (1-2504)
QY 1 CAACACATTCGTATCATTCATGATTAATGCGCAACTGCGTCAACGCGCGCTTG 60
Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929
QY 61 CCGGCGAAACCTACTGCAATTCGTTAAACCGCTTTGTTCTCCCTCAGATAAACAACAAA 120
Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnSerLys 1949
QY 121 CTGCAAAAGTTACTGGCAGACGTTAGCACACGC 153
Db 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960
RESULT 5
AAV72612
ID AAB72612 standard; Protein: 2505 AA.
XX
AC AAB72612;

RESULT 4
AAB72610
ID AAB72610 standard; Protein: 2504 AA.
XX
AC AAB72610;
XX
DT 04-MAY-2001 (first entry)
XX
DE Photorhabdus tcbA toxin.
XX
KW TcBa; TcBa; insect toxin; plant; insect resistance.
XX
OS Photorhabdus sp.
XX
PN WO200111029-A1.
XX
PD 15-FEB-2001.
XX
PF 11-AUG-2000; 2000WO-US22237.
XX
PR 11-AUG-1999; 99US-0148356.
XX
PA (DOWC) DOW AGROSCIENCES LLC.
XX
PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
PI Sukhapinda K, Merlo AO;
XX
XX WPI; 2001-191536/19.
DR N-PSDB; AAF58779.
XX
XX Novel polynucleotide sequence encoding insect toxins, useful for
XX producing transgenic plants having resistance to insects, especially
XX corn rootworm
XX
XX Disclosure; Page 62-72; 106pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC modified versions of the Photorhabdus TcBa and TcBa toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC Photorhabdus TcBa protein.
XX
SQ Sequence 2504 AA;
Alignment Scores:
Pred. No.: 2,5e-09 Length: 2504
Score: 126.50 Matches: 28
Percent Similarity: 64.71% Conservative: 5
Best Local Similarity: 54.90% Mismatches: 17
Query Match: 47.74% Indels: 1
DB: 22 Gaps: 1
US-09-856-221-2 (1-154) x AAB72610 (1-2504)
QY 1 CAACACATTCGTATCATTCATGATTAATGCGCAACTGCGTCAACGCGCGCTTG 60
Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929
QY 61 CCGGCGAAACCTACTGCAATTCGTTAAACCGCTTTGTTCTCCCTCAGATAAACAACAAA 120
Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnSerLys 1949
QY 121 CTGCAAAAGTTACTGGCAGACGTTAGCACACGC 153
Db 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960
RESULT 5
AAB72612
ID AAB72612 standard; Protein: 2505 AA.
XX
AC AAB72612;

XX 04-MAY-2001 (first entry)
XX Modified Photorhabdus tcbA toxin.
XX TcdA; TcbA; insect toxin; plant; insect resistance.
XX Photorhabdus sp.
OS Synthetic.
OS WO200111029-A1.
XX 15-FEB-2001.
XX 11-AUG-2000; 2000WO-US22237.
XX 11-AUG-1999; 99US-0148356.
XX (DOWC) DOW AGROSCIENCES LLC.
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
PI Sukhapinda K, Merlo AO;
XX WPI; 2001-191536/19.
DR N-PSDB; AAF58781.
XX Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm
XX Claim 1; Page 83-93; 106pp; English.
XX The present invention provides the protein and coding sequences of
CC modified versions of the Photorhabdus TcdA and TcbA toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC modified Photorhabdus TcbA protein.
XX Sequence 2505 AA;
SQ
Alignment Scores:
Pred. No.: 2,5e-09 Length: 2505
Score: 126.50 Matches: 28
Percent Similarity: 64.71% Conservative: 5
Best Local Similarity: 54.90% Mismatches: 17
Query Match: 47.74% Indels: 1
DB: 22 Gaps: 1
US-09-856-221-2 (1-154) x AAB72612 (1-2505)
QY 1 CAACACATTCGTGATCATTTATCAGCATAAATGCGGCAACTGGTCAACGCGCGCTTG 60
Db 1912 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1930
QY 61 CCGCGCAACACTACTGCAAAATTCGTTAAACCGCTTGTTCCTCTCAGATAAACAACAAAA 120
Db 1931 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLys 1950
QY 121 CTGCAAAATTTACTGGCAGAGCTTAGCACACGC 153
Db 1951 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1961
RESULT 6
AAW17899
ID AAW17899 standard; Protein; 2516 AA.
XX
AC AAW17899;
XX
DT 29-JAN-1998 (first entry)
XX
DE Photorhabdus luminescens insect toxin TcdA.
XX

KW Insecticide; insect; toxin; pest control; biological control;
KW Photorhabdus luminescens; TcdA; Southern corn rootworm;
KW Colorado potato beetle; Western corn rootworm; meal worm;
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
KW Diptera, Dictyoptera; Acarina; Homoptera.
XX
OS Photorhabdus luminescens strain W-14 (ATCC 55397).
XX
XX Location/Qualifiers
FH Protein 1..2516
FT /label= TcdA
FT Protein 89..1937
FT /label= TcdAii
FT Peptide 89..100
FT /note= "S2 N-terminus (Claim 30)"
FT Peptide 284..299
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 554..563
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 1080..1092
FT /note= "isolated N-terminal peptide (Claim 30)"
FT Peptide 1385..1400
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 1478..1497
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 1620..1642
FT /note= "claimed peptide (Claim 30)"
FT Peptide 1938..1948
FT /note= "N-terminal peptide (Claim 30)"
FT Protein 1938..2516
FT /label= TcdAiii
FT Peptide 2327..2345
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 2398..2408
FT /note= "tryptic peptide (Claim 30)"
XX WO9717432-A1.
XX
XX 15-MAY-1997.
XX
XX 06-NOV-1996; 96WO-US18003.
XX
XX 28-AUG-1996; 96US-0705484.
XX 06-NOV-1995; 95US-0007255.
XX 28-FEB-1996; 96US-0608423.
XX
XX (WISC) WISCONSIN ALUMNI RES FOUND.
XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI Strickland JA;
XX WPI; 1997-281022/25.
XX N-PSDB; AAT68843.
XX
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
PT can be genetically engineered into insect larvae food and plants for
PT insect control
XX
XX Claim 34; Page 185-194; 276pp; English.
XX
XX This polypeptide comprises the 282.9 kDa TcdA insecticidal toxin
CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
CC from a genomic DNA clone (AAT68843) and includes N-terminal and
CC tryptic peptide sequences obtained from the isolated protein. The
CC 282.9 kDa insect toxin is proteolytically processed into TcdAii and
CC TcdAiii components (see AAW17900 and AAW18301). Claimed toxins of P.
CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can
CC be produced by recombinant DNA methods and applied to, or genetically
CC engineered into, insect larvae food and plants for insect control.

CC The toxins are particularly effective against Southern corn rootworm,
 CC Colorado potato beetle, Western corn rootworm, meal worm, boll weevil
 CC and turf grub (Coleoptera), beet armyworm, black cutworm, cabbage
 CC looper, codling moth, corn earworm, European corn borer, tobacco
 CC hornworm and tobacco budworm (Lepidoptera), and are also active
 CC against insects of the orders Hymenoptera, Diptera, Dictyoptera,
 CC Acarina and Homoptera. (All claimed).

XX SQ Sequence 2516 AA;

Alignment Scores:
 Pred. No.: 3.69e-08 Length: 2516
 Score: 118.50 Matches: 26
 Percent Similarity: 66.04% Conservative: 9
 Best Local Similarity: 49.06% Mismatches: 15
 Query Match: 44.72% Indels: 3
 DB: 18 Gaps: 1

US-09-856-221-2 (1-154) x AAW17899 (1-2516)

QY 4 ACCATTGCGTATCATTCAGCATATAATCGCGCAACTCGCTCAACGGCGGCTTGCGG 63
 Db 1914 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1933

QY 64 GCGAAA-----CGTACTGCAAAATCGTTAACCGCTTTGTTCTCTCAGATAAAC 114
 Db 1934 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1953

QY 115 AAAAAGTCCAAAGTACTGCGACAGCTTAGCACACGC 153

Db 1954 GluValMetCAsnTyrTrpGlnThrLeuAlaGlnArg 1966

RESULT 7

AAW56572

ID AAW56572 standard; Protein; 2516 AA.

XX AC AAW56572;

XX DT 07-AUG-1998 (first entry)

XX DE Toxin TcdA, encoded by the tcdA gene from genomic region tcd.

XX KW Photobhabdus luminescens W-14; nematode; symbiotic;
 XX Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW tobacco hornworm; budworm.

XX OS Photobhabdus luminescens.

XX PN WO9808932-Al.

XX PD 05-MAR-1998.

XX PF 05-MAY-1997; 97WO-US07657.

XX PR 06-NOV-1996; 96WO-US18003.

XX PR 28-AUG-1996; 96US-0705484.

XX PR 06-NOV-1996; 96US-0743699.

XX PA (DOWC) DOWELANCO.

XX PA (WISC) WISCONSIN ALUMNI RES FOUND.

XX PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;

XX PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;

XX PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;

XX PI Strickland JA, Sukhapinda K;

XX DR WPI; 1998-179427/16.

XX DR N-PSDB; AAV29928.

XX DR

PT Isolated toxins from Photobhabdus luminescens strains - useful for
 PT control of insect pests

XX PS Claim 34; Pages 224-231; 321pp; English.

XX CC The present sequence represents a protein named TcdA of the bacterium
 CC Photobhabdus luminescens (W-14). This is a symbiotic bacterium of the
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
 CC produced from these regions that are associated with insecticidal
 CC activity. The native toxins are secreted proteins. The proteins are
 CC toxic to insects upon exposure and especially when ingested. The
 CC nucleic acid sequence can be used to produce transgenic plants,
 CC baculoviruses or microbial hosts for toxin production. They can be used
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 CC codling moth, corn earworm, European corn borer or tobacco hornworm
 CC or budworm.

XX SQ Sequence 2516 AA;

Alignment Scores:

Pred. No.: 3.69e-08 Length: 2516
 Score: 118.50 Matches: 26
 Percent Similarity: 66.04% Conservative: 9
 Best Local Similarity: 49.06% Mismatches: 15
 Query Match: 44.72% Indels: 3
 DB: 19 Gaps: 1

US-09-856-221-2 (1-154) x AAW56572 (1-2516)

QY 4 ACCATTGCGTATCATTCAGCATATAATCGCGCAACTCGCTCAACGGCGGCTTGCGG 63

Db 1914 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1933

QY 64 GCGAAA-----CGTACTGCAAAATCGTTAACCGCTTTGTTCTCTCAGATAAAC 114

Db 1934 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1953

QY 115 AAAAAGTCCAAAGTACTGCGACAGCTTAGCACACGC 153

Db 1954 GluValMetCAsnTyrTrpGlnThrLeuAlaGlnArg 1966

RESULT 8

AAW72609

ID AAW72609 standard; Protein; 2516 AA.

XX AC AAW72609;

XX DT 04-MAY-2001 (first entry)

XX DE Photobhabdus tcdA toxin.

XX KW TcdA; TcBA; insect toxin; plant; insect resistance.

XX OS Photobhabdus sp.

XX PN WO200111029-Al.

XX PD 15-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US22237.

XX PR 11-AUG-1999; 99US-0148356.

XX PA (DOWC) DOW AGROSCIENCES LLC.

XX PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

XX PI Sukhapinda K, Merlo AO;

XX DR WPI; 2001-191536/19.

XX DR

DR N-PSDB; AAF58778.
XX Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm -
XX
PS Disclosure; Page 51-61; 106pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC modified versions of the Photobabidus Tcda and Tcda toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC modified Photobabidus Tcda protein.
XX
SQ Sequence 2516 AA;
Alignment Scores: 3.69e-08 Length: 2516
Pred. No.: 118.50 Matches: 26
Score: 118.50
Percent Similarity: 66.04% Conservative: 9
Best Local Similarity: 49.06% Mismatches: 15
Query Match: 44.72% Indels: 3
DB: 22 Gaps: 1
US-09-856-221-2 (1-154) x AAB72609 (1-2516)
QY 4 ACCATTCGTGATCATATCAGCATAAAATGCGCAACTGCGTCAACGCGCGCTTGC CG 63
||| : : : : : |||
Db 1914 ThrThrGlnAsnAlaHisaspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1933
QY 64 GCGAAA-----CGTACTGCAAAATTCGTTAACCGCTTTGCTTCTCCTCAGATAAAC 114
||| : : : : : |||
Db 1934 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1953
QY 115 AAAAAGTCAAGTTACTGCGCAGCGTTAGCACAACGC 153
: : : : :
Db 1954 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1966
RESULT 9
AAB72611
ID AAB72611 standard; Protein; 2517 AA.
XX
AC AAB72611;
XX
DT 04-MAY-2001 (first entry)
XX
DE Modified Photobabidus tcda toxin.
XX
KW Tcda; Tcda; insect toxin; plant; insect resistance.
XX
OS Photobabidus sp.
OS Synthetic.
XX
PN WO200111029-A1.
XX
PD 15-FEB-2001.
XX
PF 11-AUG-2000; 2000WO-US22237.
XX
PR 11-AUG-1999; 99US-0148356.
XX
PA (DOWC) DOW AGROSCIENCES LLC.
XX
PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
PI Sukhapinda K, Merlo AO;
XX
DR WPI; 2001-191536/19.
DR N-PSDB; AAF58780.
XX
XX Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm -
XX

XX Claim 1; Page 72-83; 106pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC modified versions of the Photobabidus Tcda and Tcda toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC modified Photobabidus Tcda protein.
XX
SQ Sequence 2517 AA;
Alignment Scores: 3.69e-08 Length: 2517
Pred. No.: 118.50 Matches: 26
Score: 118.50
Percent Similarity: 66.04% Conservative: 9
Best Local Similarity: 49.06% Mismatches: 15
Query Match: 44.72% Indels: 3
DB: 22 Gaps: 1
US-09-856-221-2 (1-154) x AAB72611 (1-2517)
QY 4 ACCATTCGTGATCATATCAGCATAAAATGCGCAACTGCGTCAACGCGCGCTTGC CG 63
||| : : : : : |||
Db 1915 ThrThrGlnAsnAlaHisaspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1934
QY 64 GCGAAA-----CGTACTGCAAAATTCGTTAACCGCTTTGCTTCTCCTCAGATAAAC 114
||| : : : : : |||
Db 1935 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1954
QY 115 AAAAAGTCAAGTTACTGCGCAGCGTTAGCACAACGC 153
: : : : :
Db 1955 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1967
RESULT 10
AAB72614
ID AAB72614 standard; Protein; 2537 AA.
XX
AC AAB72614;
XX
DT 04-MAY-2001 (first entry)
XX
DE Tcda toxin-zein ER signal peptide fusion protein.
XX
KW Tcda; Tcda; insect toxin; plant; insect resistance.
XX
OS Chimeric - Photobabidus sp.
OS Chimeric - Zea mays.
XX
PN WO200111029-A1.
XX
PD 15-FEB-2001.
XX
PF 11-AUG-2000; 2000WO-US22237.
XX
PR 11-AUG-1999; 99US-0148356.
XX
PA (DOWC) DOW AGROSCIENCES LLC.
XX
PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
PI Sukhapinda K, Merlo AO;
XX
DR WPI; 2001-191536/19.
DR N-PSDB; AAF58783.
XX
XX Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm -
XX
PS Example 1; Page 94-104; 106pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC modified versions of the Photobabidus Tcda and Tcda toxins. These are

This sequence represents the hph2 gene encoded insecticidal toxin from *Photographus luminescens*. It is one of three insecticidal toxins (AAV33728-V33730) encoded by open reading frames (orfs) in a 38kb fragment of *P. luminescens* DNA (AAZ06831). The hph2 gene was identified using probe #2 which was amplified from *P. luminescens* genomic DNA using PCR primers AAZ06829-206830. *P. luminescens* is a member of the

PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA;
 XX WPI; 1997-281022/25.
 DR N-PSDB; AAT68847.
 XX
 XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
 PT can be genetically engineered into insect larvae food and plants for
 PT insect control
 PT
 XX Claim 34; Page 229-231; 276pp; English.
 XX
 CC This polypeptide comprises the 62.9 kDa TcbAIII insecticidal toxin
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
 CC from a genomic DNA clone (AAT68847). TcbAIII is a proteolytic
 CC cleavage product of TcBA (see AAW17871). Claimed toxins of P.
 CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can
 CC be produced by recombinant DNA methods and applied to, or genetically
 CC engineered into, insect larvae food and plants for insect control.
 CC The toxins are particularly effective against Southern corn rootworm,
 CC Colorado potato beetle, Western corn rootworm, meal worm, boll
 CC weevil and turf grub (Coleoptera), beet armyworm, black cutworm,
 CC cabbage looper, codling moth, corn earworm, European corn borer,
 CC tobacco hornworm and tobacco budworm (Lepidoptera), and are also
 CC active against insects of the orders Hymenoptera, Diptera,
 CC Dictyoptera, Acarina and Homoptera. (All claimed).
 XX
 SQ Sequence 573 AA;
 Alignment Scores:
 Pred. No.: 1.66e-07 Length: 573
 Score: 113.00 Matches: 22
 Percent Similarity: 88.89% Conservative: 2
 Best Local Similarity: 81.48% Mismatches: 3
 Query Match: 42.64% Indels: 0
 DB: 18 Gaps: 0
 05-09-856-221-2 (1-154) x AAW18303 (1-573)
 QY 73 ACTGCAAAATTCGTTAACCGCTTTGTCCTTCAGATAAACAACAACTGCAAGTTAC 132
 Db 3 ThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLysLeuLysGlyTyr 22
 QY 133 TGGCAGACGTTAGCACACGC 153
 Db 23 TrpArgThrLeuAlaGlnArg 29
 RESULT 13
 AAW56559
 ID AAW56559 standard; Protein; 573 AA.
 AC AAW56559;
 XX
 XX 07-AUG-1998 (first entry)
 XX
 XX Toxin TcbAIII, encoded by the tcbA gene from genomic region tcb.
 XX
 KW Photorhabdus luminescens W-14; nematode; symbiotic;
 KW Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW Tobacco hornworm; budworm.
 XX
 OS Photorhabdus luminescens.
 XX
 XX W09808932-AL.
 XX
 XX 05-MAR-1998.
 PD
 XX 05-MAY-1997; 97WO-US07657.
 PF
 XX

PR 06-NOV-1996; 96WO-US18003.
 PR 28-AUG-1996; 96US-0705484.
 XX 06-NOV-1996; 96US-0743699.
 XX
 PA (DOWC) DOWELANCO.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PA Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA, Sukhapinda K;
 XX
 DR WPI; 1998-179427/16.
 DR N-PSDB; AAV29987.
 XX
 PT Isolated toxins from Photorhabdus luminescens strains - useful for
 PT control of insect pests
 PT
 XX Claim 34; Pages 261-263; 321pp; English.
 XX
 CC The present sequence represents a protein named TcbAIII of the bacterium
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
 CC produced from these regions that are associated with insecticidal
 CC activity. The native toxins are secreted proteins. The proteins are
 CC toxic to insects upon exposure and especially when ingested. The
 CC nucleic acid sequence can be used to produce transgenic plants. The
 CC baculoviruses or microbial hosts for toxin production. They can be used
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 CC codling moth, corn earworm, European corn borer or tobacco hornworm
 CC or budworm.
 XX
 SQ Sequence 573 AA;
 Alignment Scores:
 Pred. No.: 1.66e-07 Length: 573
 Score: 113.00 Matches: 22
 Percent Similarity: 88.89% Conservative: 2
 Best Local Similarity: 81.48% Mismatches: 3
 Query Match: 42.64% Indels: 0
 DB: 19 Gaps: 0
 US-09-856-221-2 (1-154) x AAW56559 (1-573)
 QY 73 ACTGCAAAATTCGTTAACCGCTTTGTCCTTCAGATAAACAACAACTGCAAGTTAC 132
 Db 3 ThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLysLeuLysGlyTyr 22
 QY 133 TGGCAGACGTTAGCACACGC 153
 Db 23 TrpArgThrLeuAlaGlnArg 29
 RESULT 14
 AAW18301
 ID AAW18301 standard; Protein; 579 AA.
 XX
 AC AAW18301;
 XX
 XX 29-JAN-1998 (first entry)
 XX
 XX Photorhabdus luminescens insect toxin TcdAIII.
 DE
 XX
 XX Insecticide; insect; toxin; pest control; biological control;
 KW Photorhabdus luminescens; tcdA; Southern corn rootworm;
 KW Colorado potato beetle; Western corn rootworm; meal worm;
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
 KW Diptera, Dictyoptera; Acarina; Homoptera.

Db 22 TyrTrpGlnThrLeuAlaGlnArg 29

Search completed: January 15, 2003, 15:50:28
Job time : 19.2569 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:45:23 ; Search time 4.96166 Seconds
(without alignments)
1826.456 Million cell updates/sec

Title: US-09-856-221-2
Perfect score: 265
Sequence: 1 caaacattcgatcattatcagcattaaatcgcgcaactgcgtcaacgcgcg --- GCCTTG 60

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frames_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09856221/runat_15012003_153926_12521/app_query.fasta_1.1948
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221 -CGN_1_1_24=runat_15012003_153926_12521 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMBOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.5	43.6	2522	4	US-09-251-645-13
2	56.5	21.2	951	1	US-08-162-809-2
3	56	21.1	355	4	US-08-984-618-14
4	55	20.6	771	4	US-09-462-284-2
5	53	19.9	329	4	US-09-651-200-19
6	52.5	19.8	443	1	US-08-399-986B-2
7	52.5	19.8	443	2	US-08-493-754A-2
8	52.5	19.8	1048	4	US-08-887-534A-85
9	51.5	19.3	227	3	US-09-248-325-46
10	51	19.2	236	4	US-09-121-979-4
11	51	19.2	236	4	US-09-332-319-4
12	51	19.2	236	4	US-09-239-867-2

13	51	19.2	616	4	US-09-134-001C-4618	Sequence 4618, Ap
14	50.5	19.1	575	4	US-09-171-461-7	Sequence 7, Appl
15	50.5	18.9	984	2	US-08-673-789-6	Sequence 6, Appl
16	50.5	18.9	1457	3	US-08-665-259-27	Sequence 27, Appl
17	50.5	18.9	1457	3	US-08-762-500-27	Sequence 27, Appl
18	50	18.7	33	4	US-08-981-392-63	Sequence 63, Appl
19	50	18.7	66	3	US-08-908-643C-11	Sequence 11, Appl
20	50	18.7	76	3	US-08-908-643C-9	Sequence 9, Appl
21	50	18.7	92	3	US-08-908-643C-7	Sequence 7, Appl
22	50	18.7	124	3	US-08-908-643C-5	Sequence 5, Appl
23	50	18.7	257	2	US-08-637-759B-92	Sequence 92, Appl
24	50	18.7	257	3	US-08-871-355A-92	Sequence 92, Appl
25	50	18.7	257	3	US-09-201-945-92	Sequence 92, Appl
26	50	18.7	494	4	US-08-378-313-23	Sequence 23, Appl
27	50	18.7	494	4	US-08-378-313-29	Sequence 29, Appl
28	50	18.9	698	4	US-09-134-001C-3632	Sequence 3632, Ap
29	49.5	18.7	148	1	US-08-468-347-22	Sequence 22, Appl
30	49.5	18.7	148	2	US-08-467-389-22	Sequence 22, Appl
31	49.5	18.7	148	2	US-08-779-379-22	Sequence 22, Appl
32	49.5	18.7	148	2	US-08-469-219-22	Sequence 22, Appl
33	49.5	18.7	148	4	US-09-228-152-22	Sequence 22, Appl
34	49.5	18.5	339	3	US-08-758-280-1	Sequence 1, Appl
35	49.5	18.5	339	3	US-08-758-280-2	Sequence 2, Appl
36	49.5	18.5	339	3	US-08-964-614A-1	Sequence 1, Appl
37	49.5	18.5	339	3	US-08-964-614A-2	Sequence 2, Appl
38	49.5	18.7	374	1	US-08-468-847B-12	Sequence 12, Appl
39	49.5	18.7	375	2	US-08-459-101A-2	Sequence 2, Appl
40	49.5	18.5	3165	2	US-08-459-146-3	Sequence 3, Appl
41	49.5	18.5	3165	2	US-08-459-065-3	Sequence 3, Appl
42	49	18.5	329	4	US-09-011-769A-51	Sequence 51, Appl
43	49	18.5	349	4	US-09-011-769A-47	Sequence 47, Appl
44	49	18.5	349	4	US-09-011-769A-60	Sequence 60, Appl
45	49	18.5	349	4	US-09-011-769A-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1
US-09-251-645-13
; Sequence 13, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-13

Alignment Scores:
Pred. No.: 1.19e-08 Length: 2522
Score: 115.50 Matches: 26
Percent Similarity: 68.63% Conservative: 9
Best Local Similarity: 50.98% Mismatches: 15
Query Match: 43.58% Indels: 1
DB: Gaps: 1

US-09-856-221-2 (1-154) x US-09-251-645-13 (1-2522)

Oy 4 ACCATTGTCATTCATTATCAGCATAAAATCGCGCAACTGCCTCAACGCGCG --- GCCTTG 60

```

Db 1922 ThrThrGlnSerAlaHisSerSerIleValAlaLeuArgGlnSerThrProAlaLeu 1941
QY 61 CCGCGCAACGACTGCAAAATTCGTTAAACCGCTTTGTCCTTCCTCAGATAAACAATAA 120
Db 1942 LeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsnGluVal 1961
QY 121 CTCGAAGTACTGCGCAGACGCTTAGCACAACGC 153
Db 1962 MetMetAsnTyTrpGlnThrLeuAlaGlnArg 1972

RESULT 2
US-08-162-809-2
; Sequence 2, Application US/08162809
; Patent No. 5457048
; GENERAL INFORMATION:
; APPLICANT: Pasquale, Elena B.
; APPLICANT: Sajjadi, Fereydoun G.
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,809
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 951 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-162-809-2
Alignment Scores:
Pred. No.: 6.39 Length: 951
Score: 56.50 Matches: 17
Percent Similarity: 41.51% Conservative: 5
Best Local Similarity: 32.08% Mismatches: 27
Query Match: 21.16% Indels: 4
DB: 1 Gaps: 1

US-09-856-221-2 (1-154) x US-08-162-809-2 (1-951)
QY 152 CGTTGCTGAACGTCGCGCAGTAAGTTGTCAGTTT-----TGTATCTGAGGA 102
Db 143 LysCysProSerValValGln-AsnPheAlaIlePheProGluThrMetThrGlyAlaG1 162
QY 101 AGGAACAACGCGTTTAACGAATTTGTCAGTAGCTTCGCGCGCAAGCGCGGTGACGC 42
Db 162 uSerThrSerLeuValThrAlaArgGlyThrCysIleProAsnAlaGluGluValAspVa 182
QY 41 AGTTGCGCGCATTTATGCTGATGAATGATCAGCAATGG 5
Db 182 lProIleLysLeuTyrcysAsnGlyAspGlyGluTrp 194

```

```

RESULT 3
US-08-984-618-14
; Sequence 14, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomasz, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: E. coli
US-08-984-618-14
Alignment Scores:
Pred. No.: 5.61 Length: 355
Score: 56.00 Matches: 18
Percent Similarity: 37.14% Conservative: 8
Best Local Similarity: 25.71% Mismatches: 22
Query Match: 21.13% Indels: 22
DB: 4 Gaps: 2

US-09-856-221-2 (1-154) x US-08-984-618-14 (1-355)
QY 1 CAACCATTCGTGTCATCTAT-----CAGCATAAAATCGCGCAACTG 42
Db 224 GlnSerValGluGlnAlaTyrAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 243
QY 43 -----CGTCACCGCGG 54
Db 244 IleAspAspMetAlaAlaTyrAlaTyrAlaTyrAlaAspValValCysArgSerGlyAla 263
QY 55 GCCTTGGCGCGCAACGCTACTGCATAATTCGTTAAACGCTTTGTTCTTCCTCAGATAAAC 114
Db 264 LeuThrValSerGluIleAlaAlaGlyLeuProAlaLeuPheValProPheGlnHis 283
QY 115 AAAAACTGCAGAACTTACTGCGCAGACGTTA 144
Db 284 LysAspArgGlnGlnTyrTrpAsnAlaLeu 293

```


Db 196 TyrileAlaValPheIleProCysGlyHisLeuValThrCysLysGlnCysAlaGlu 215
Qy 68 AAGCTACTCAAAATCGTTAACCGCTTTGTTC 100
Db 216 AlaVal-----AspArgCysPrometCysSer 224
RESULT 13
US-09-134-001C-4618
; Sequence 4618, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4618
; LENGTH: 616
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4618

Alignment Scores:
Pred. No.: 37.4 Length: 616
Score: 51.00 Matches: 10
Percent Similarity: 64.00% Conservative: 6
Best Local Similarity: 40.00% Mismatches: 9
Query Match: 19.25% Indels: 0
DB: 4 Gaps: 0

US-09-856-221-2 (1-154) x US-09-134-001C-4618 (1-616)

Qy 22 CAGCATAAAATCGGCAACTCGTCAACGGCGGCTTGGCGGCAAACTACTGCAAT 81
Db 80 GluHisLysValIleGluHisGlnAlaValGlnIleAlaLysArgAlaAlaAsn 99
Qy 82 TCGTTAACCGCTTG 96
Db 100 GluAspThrIleIle 104

RESULT 14
US-09-171-461-7
; Sequence 7, Application US/09171461
; Patent No. 6335016
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotton, Matthew
; APPLICANT: Chiocci, Susanna
; APPLICANT: Kurzbaner, Robert
; APPLICANT: Schaefner, Gotthold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800000
; CURRENT APPLICATION NUMBER: US/09/171,461
; CURRENT FILING DATE: 1999-01-12
; EARLIER APPLICATION NUMBER: PCT/EP97/01944
; EARLIER FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 575
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 13316..15043 /gene: L1 /product: L1 IIfa
US-09-171-461-7

Alignment Scores:

Pred. No.: 43.6 Length: 575
Score: 50.50 Matches: 17
Percent Similarity: 56.52% Conservative: 9
Best Local Similarity: 36.96% Mismatches: 11
Query Match: 19.06% Indels: 9
DB: 4 Gaps: 3

US-09-856-221-2 (1-154) x US-09-171-461-7 (1-575)

Qy 22 CAGCATAAAATCGGCAACTCGCTCAA-----CGCGGCGCTTGGCGGAAAGTACT 75
Db 316 LysTyrLeuMetArgGlnLeuArgGlnAlaLeuLysAspGlyValProAlaAspIleSer 335
Qy 76 GCAAAATCGTTAAACGGCTTGTTCCTTCCT-----CAGATAAACAAA----- 117
Db 336 IleSerThrMetThrGlnTyrLeuAspProArgLeuTyrGlnThrAsnLysValPheVal 355
Qy 118 ---AACTGCAAAATTAC 132
Db 356 GluLysLeuGlnAsnTyr 361

RESULT 15

US-08-673-789-6
; Sequence 6, Application US/08673789
; Patent No. 5814479

GENERAL INFORMATION:
; APPLICANT: ZHOU, RENPING; SCHULZ, NICHOLAS, KROMER, LAWRENCE, F.; VANDE WOUDE,
; APPLICANT: GEORGE, F.
; TITLE OF INVENTION: BSK RECEPTOR LIKE
; TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR
; TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/673,789
; FILING DATE:
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/177,812
; FILING DATE: 04-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
US-08-673-789-6

Alignment Scores:
Pred. No.: 51.4 Length: 984
Score: 50.50 Matches: 16
Percent Similarity: 39.62% Conservative: 5

```

Best Local Similarity: 30.19%
Query Match: 18.91%
DB: 2
Mismatches: 28
Indels: 4
Gaps: 1

US-09-856-221-2 (1-154) x US-08-673-789-6 (1-984)

Qy 152 CGTTGCTAAGCTGCTGCCAGTAACTTTGCGAGTTT-----TCTTTATCTGAGGA 102
Db 195 LysCysProSerIleValGln-AsnPheAlaValPheProGluThrMetThrGlyAlaG1 214
Qy 101 AGGAACAAGCGGTAAAGAAATTTGCGAGTACGTTCGCGCAAGCGCGCGTTGACGC 42
Db 214 uSerThrSerLeuValIleAlaArgGlyThrCysIleProAsnAlaGluValAspVa 234
Qy 41 AGTGGCGCATTTTATGCTGATATGATCAGGAATGG 5
Db 234 lProileLysLeuTyrcysAsnGlyAspGlyGluTrp 246

```

Search completed: January 15, 2003, 15:57:44
Job time : 6.96166 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:46:43 ; Search time 2.88907 Seconds
(without alignments)
2119.222 Million cell updates/sec

Title: US-09-856-221-2
Perfect score: 265
Sequence: 1 caaacattcgtgacattca.....gcagacgttagcacacagccc 154

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=rlp
-Q=/cgn2_1/USPTO_Spool/US09856221/runat_15012003_153926_12555/app_query.fasta_1.1948
-DB=Published_Applications_AA -QPMF=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09856221 -CGN_1_14 -runat_15012003_153926_12555
-NCPU=6 -ICPU=3 -NO_XLUPY -NO_MMUP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126.5	47.7	2504	10	US-09-817-514A-8
2	118.5	44.7	2516	10	US-09-817-514A-2
3	57	21.3	322	9	US-09-966-346-20
4	57	21.3	322	9	US-09-966-345-20

c	5	57	21.3	322	9	US-09-965-212-20	Sequence 20, Appl
c	6	57	21.3	322	9	US-10-174-590-172	Sequence 172, App
c	7	57	21.3	322	9	US-10-176-758-172	Sequence 172, App
c	8	57	21.3	322	12	US-10-052-586-172	Sequence 172, App
c	9	56	21.1	355	10	US-09-741-669-454	Sequence 454, App
c	10	56	21.1	355	10	US-09-815-242-10031	Sequence 10031, A
c	11	54.5	20.4	1232	10	US-09-801-574-46	Sequence 46, Appl
c	12	53.5	20.0	302	10	US-09-903-814-14	Sequence 14, Appl
c	13	53	20.0	215	9	US-09-764-868-1004	Sequence 1004, Ap
c	14	53	19.9	312	10	US-09-801-368-438	Sequence 438, App
c	15	53	20.0	355	10	US-09-815-242-14078	Sequence 14078, A
c	16	52.5	19.8	1048	10	US-09-741-669-409	Sequence 409, App
c	17	52.5	19.8	1048	10	US-09-815-242-10062	Sequence 10062, A
c	18	52	19.6	566	10	US-09-801-368-418	Sequence 418, App
c	19	52	19.6	780	9	US-09-738-626-4706	Sequence 4706, Ap
c	20	51.5	19.4	324	10	US-09-815-242-13895	Sequence 13895, A
c	21	51.5	19.3	333	10	US-09-815-242-10116	Sequence 10116, A
c	22	51.5	19.4	1046	9	US-10-078-770-186	Sequence 186, App
c	23	51	19.2	334	9	US-09-738-626-5255	Sequence 5255, Ap
c	24	51	19.2	626	9	US-09-738-626-5415	Sequence 5415, Ap
c	25	50.5	19.1	575	10	US-09-970-711-7	Sequence 7, Appli
c	26	50.5	18.9	2001	9	US-10-072-621-8	Sequence 8, Appli
c	27	50.5	18.9	2436	10	US-09-795-693-8	Sequence 8, Appli
c	28	50	18.7	33	10	US-09-908-322-63	Sequence 63, Appl
c	29	50	18.9	206	10	US-09-815-242-5290	Sequence 5290, Ap
c	30	50	18.9	207	10	US-09-815-242-12240	Sequence 12240, A
c	31	50	18.9	207	10	US-09-815-242-12758	Sequence 12758, A
c	32	50	18.9	419	10	US-09-894-018-123	Sequence 123, App
c	33	50	18.5	647	10	US-09-725-735A-18	Sequence 23, Appl
c	34	49.5	18.7	62	10	US-09-863-693-23	Sequence 23, Appl
c	35	49.5	18.7	148	10	US-09-864-761-41723	Sequence 41723, A
c	36	49.5	18.7	374	10	US-09-864-761-46739	Sequence 46739, A
c	37	49.5	18.7	374	10	US-09-853-6258-12	Sequence 12, Appl
c	38	49.5	18.5	598	10	US-09-875-811-12	Sequence 8, Appli
c	39	49.5	18.5	605	10	US-09-875-811-4	Sequence 4, Appli
c	40	49.5	18.5	621	10	US-09-875-811-10	Sequence 10, Appl
c	41	49.5	18.5	656	10	US-09-942-446-2	Sequence 2, Appli
c	42	49.5	18.5	656	10	US-09-875-811-6	Sequence 6, Appli
c	43	49.5	18.5	663	10	US-09-875-811-2	Sequence 2, Appli
c	44	49.5	18.5	679	10	US-09-858-525A-10	Sequence 10, Appl
c	45	49.5	18.7	836	9	US-09-858-525A-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: french-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Alignment Scores:
Pred. No.: 2.47e-10
Score: 126.50
Percent Similarity: 64.71%
Best Local Similarity: 54.90%
Matches: 28
Conservative: 5
Mismatches: 17

```
Query Match: 47.74% Indels: 1
DB: 10 Gaps: 1
US-09-856-221-2 (1-154) x US-09-817-514A-8 (1-2504)
QY 1 CAAACCATTCGTGATCATTAATGCGGCAATGCGTCAACGCGGCGCTTG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1911 GlnGlnValArgGlnGlnValThrGlnLeuArg---LeuAsnSerArgValLysThr 1929
QY 61 CCGGCGCAAGACTGCGAAATCTTAAACGCTTGTTCTCTCCACAGATAAACA 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLys 1949
QY 121 CTGCAAGATTACTGGCAGACCTTAGCACAAACGC 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1950 LeuLysGlyTyrTyrArgThrLeuAlaGlnArg 1960
RESULT 2
US-09-817-514A-2
; Sequence 2, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: firench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2
Alignment Scores:
Pred. No.: 3 77e-09 Length: 2516
Score: 118.50 Matches: 26
Percent Similarity: 66.04% Conservative: 9
Best Local Similarity: 49.06% Mismatches: 15
Query Match: 44.72% Indels: 3
DB: 10 Gaps: 1
US-09-856-221-2 (1-154) x US-09-817-514A-2 (1-2516)
QY 4 ACCATTGCGATCATTAATGCGGCAATGCGTCAACGCGGCGCTTGCGG 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1914 ThrThrGlnAsnAlaHisAspSerAlaValAlaLeuArgGlnAsnIleProThrPro 1933
QY 64 GCGAAA-----CGTACTGAAATTCGTTAACCGCTTTGTTCTCTCCACAGATAAAC 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1934 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1953
QY 115 AAAAACTGCAAGTTACTGGCAGACCTTAGCACAAACGC 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1954 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1966
RESULT 3
US-09-966-546-20
; Sequence 20, Application US/09966546
; Patent No. US20020168716A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20020168716A1 Human Proteins and Polynucleotides Encoding
; FILE REFERENCE: Cura-46 (15966-546)
```

```
; CURRENT APPLICATION NUMBER: US/09/966,546
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/544,511
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 20
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-546-20
Alignment Scores:
Pred. No.: 2 76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: 9 Gaps: 0
US-09-856-221-2 (1-154) x US-09-966-546-20 (1-322)
QY 145 CTAACGTCGCCAGTAACCTTTCAGCTTTTGTATCTGAGGAAGAAACAAAGCGTTA 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 IleThrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyrProThr 122
QY 85 ACGAATTTGCAGTACGTTTCGCGGCAAGCGCGCTTGACGAGTTGCCGATTTTAT 26
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaAlaThrPhe 142
RESULT 4
US-09-966-545-20
; Sequence 20, Application US/09966545
; Patent No. US20020172999A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20020172999A1 Human Proteins and Polynucleotides Encod
; FILE REFERENCE: Cura-46 (15966-546)
; CURRENT APPLICATION NUMBER: US/09/966,545
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/544,511
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 20
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-545-20
Alignment Scores:
Pred. No.: 2 76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: 9 Gaps: 0
US-09-856-221-2 (1-154) x US-09-966-545-20 (1-322)
QY 145 CTAACGTCGCCAGTAACCTTTCAGCTTTTGTATCTGAGGAAGAAACAAAGCGTTA 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 IleThrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyrProThr 122
QY 85 ACGAATTTGCAGTACGTTTCGCGGCAAGCGCGCTTGACGAGTTGCCGATTTTAT 26
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaAlaThrPhe 142
RESULT 5
US-09-965-212-20
```

```

; Sequence 20, Application US/09965212
; Publication No. US20030003462A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Shmukets, Richard A.
; TITLE OF INVENTION: Human Proteins and Polynucleotides Encoding
; FILE REFERENCE: Cura-46 (15966-546)
; CURRENT APPLICATION NUMBER: US/09/965.212
; CURRENT FILING DATE: 2001-09-26
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: USSN 60/128.514
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-212-20

Alignment Scores:
Pred. No.: 2.76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservatives: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: Gaps: 0

US-09-856-221-2 (1-154) x US-09-965-212-20 (1-322)
QY 145 CTAACGTCGCAGTACCTTTCAGTCTTTTGTATCTGAGGAGCAACAAAGCGTTA 86
Db 103 IlethrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIlelleTyrProThr 122
QY 85 ACGAATTTCAGTACGTTTCGCGGCAAGCGCGTTCAGCGAGTTGCGCGCATTTTAT 26
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaAlaThrPhe 142

RESULT 6
; Sequence 172, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174.590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 172
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-172

Alignment Scores:
Pred. No.: 2.76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservatives: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: Gaps: 0

US-09-856-221-2 (1-154) x US-10-174-590-172 (1-322)
QY 145 CTAACGTCGCAGTACCTTTCAGTCTTTTGTATCTGAGGAGCAACAAAGCGTTA 86
Db 103 IlethrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIlelleTyrProThr 122
QY 85 ACGAATTTCAGTACGTTTCGCGGCAAGCGCGTTCAGCGAGTTGCGCGCATTTTAT 26
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaAlaThrPhe 142

RESULT 7
; Sequence 172, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176.758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 172
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-172

Alignment Scores:
Pred. No.: 2.76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservatives: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: Gaps: 0

US-09-856-221-2 (1-154) x US-10-176-758-172 (1-322)
QY 145 CTAACGTCGCAGTACCTTTCAGTCTTTTGTATCTGAGGAGCAACAAAGCGTTA 86
Db 103 IlethrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIlelleTyrProThr 122
QY 85 ACGAATTTCAGTACGTTTCGCGGCAAGCGCGTTCAGCGAGTTGCGCGCATTTTAT 26
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaAlaThrPhe 142

RESULT 8
; Sequence 172, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.

```

APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05

;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088722
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088740
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088811
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088825
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088863
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089090
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089908

Alignment Scores:
Pred. No.: 2,76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: 12 Gaps: 0

US-09-856-221-2 (1-154) x US-10-052-586-172 (1-322)

Qy 145 CTACCTGTCGACGTAACCTTTCGAGTTTGTATCTGAGGAGCAACGCGTTA 86
Db 103 IleThrPheAlaCysTyTyrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyProThr 122
Qy 85 ACGAATTGTCAGTACGTTTCGCGGCAAGCGCGGTGACGAGTTCGCGCATTTTAT 26
Db 123 ThrTyTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaAlaThrPhe 142

RESULT 9
US-09-741-669-454
; Sequence 454, Application US/09741669
; Patent No. US2002022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for

;; TITLE OF INVENTION: proliferation of E. coli
;; FILE REFERENCE: ELITRA.009A
;; CURRENT APPLICATION NUMBER: US/09/741,669
;; CURRENT FILING DATE: 2000-12-19
;; PRIOR APPLICATION NUMBER: US 60/173005
;; PRIOR FILING DATE: 1999-12-23
;; NUMBER OF SEQ ID NOS: 481
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 454
;; LENGTH: 355
;; TYPE: PRT
;; ORGANISM: Escherichia coli
US-09-741-669-454

Alignment Scores:
Pred. No.: 3,98 Length: 355
Score: 56.00 Matches: 18
Percent Similarity: 37.14% Conservative: 8
Best Local Similarity: 25.71% Mismatches: 22
Query Match: 21.13% Indels: 22
DB: 10 Gaps: 2

US-09-856-221-2 (1-154) x US-09-741-669-454 (1-355)

Qy 1 CAACCATTCGTGATCATTTAT-----CAGCATAAATGCGCGCAACTG 42
Db 224 GlnSerValGluGlnAlaTyTyrAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 243
Qy 43 -----CGTCAACGCGG 54
Db 244 IleAspAspMetAlaAlaAlaTyTyrAlaTrpAlaAspValValValCysArgSerGlyAla 263
Qy 55 GCCTTGGCGCGCAACGTACTGCAAAATTCGTTAACCGCTTGTCTTCTCAGATAAAC 114
Db 264 LeuThrValSerGluIleAlaAlaAlaGlyLeuProAlaLeuPheValProPheGlnHis 283
Qy 115 AAAAACTGCAAGTTACTGCGAGCGTTA 144
Db 284 LysAspArgGlnGlnTyTyrTrpAsnAlaLeu 293

RESULT 10
US-09-815-242-10031
; Sequence 10031, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

US-09-856-221-2 (1-154) x US-09-801-574-46 (1-1232)

Oy	154	GGCGTGTGCTAACGCTGCCAGTAACATTTGCAGTGTCTTTCTTATCTCGAGGAGCAACA	95
Db	147	GlyPheIleHisArgSerLeuSerTyrAlaValHisIleValSer---	165
Oy	94	AAGCGGTTAACGAATTTCAGTACGCTTTCGCCGCAAGG	56
Db	166	AlaArgLeuThrAspLeuGlnTyrLeuThrGluSerGln	178

US-09-856-221-2 (1-154) x US-09-764-868-1004 (1-215)

QY 58 TTGCGCGGAAGTACTGCAAAATCGTTAACGGCTTTCTCTCTCCAGATAACAAA 117
||||| |||:|||| |||:|||| |||||| |||
Db 183 LeuProGlyLysThrPheAsnMetAspArgAspPheLeuGluLysArgLysLys 202
||| |||:|||| |||:|||| |||:|||| |||
QY 118 AAATGCAAGTTACTGCGCAGAGCTTA 144
||| |||:|||| |||:|||| |||
Db 203 AspLeuAsnAlaTyrLeuGlnLeu 211
||| |||:|||| |||:|||| |||

RESULT 14

US-09-801-368-438
; Sequence 438, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 438
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-438

Alignment Scores:
Pred. No.: 10.7 Length: 312
Score: 53.00 Matches: 12
Percent Similarity: 56.67% Conservative: 5
Best Local Similarity: 40.00% Mismatches: 13
Query Match: 19.85% Indels: 0
DB: 10 Gaps: 0

US-09-856-221-2 (1-154) x US-09-801-368-438 (1-312)

QY 133 AGTAACTTTCAGTTTTTTTATCTCAGGAAGCAACAAAGCGTTTAACGAATTTGCAG 74
||||| |||:|||| |||:|||| |||:|||| |||
Db 273 SerAsnPheLysIlePheThrLeuProGluAspPheLysThrIleSerAsnLeuSer 292
||| |||:|||| |||:|||| |||:|||| |||
QY 73 TACGTTTCGCGCAAGCGCGCGCTTCAC 44
||| |||:|||| |||:|||| |||
Db 293 LysValHisGlyThrLysArgValValAsp 302
||| |||:|||| |||:|||| |||

RESULT 15

US-09-815-242-14078
; Sequence 14078, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14078
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-14078

Alignment Scores:
Pred. No.: 11 Length: 355
Score: 53.00 Matches: 15
Percent Similarity: 38.60% Conservative: 7
Best Local Similarity: 26.32% Mismatches: 19
Query Match: 20.00% Indels: 16
DB: 10 Gaps: 1

US-09-856-221-2 (1-154) x US-09-815-242-14078 (1-355)

QY 22 CAGCATAAATGCGCAACTG----- 42
||||| |||:|||| |||:|||| |||:|||| |||
Db 237 GlnHisLysValThrGluPheIleAspGlyMetAlaAlaTyrAlaTrpAlaAspVal 256
||| |||:|||| |||:|||| |||:|||| |||
QY 43 -----CGTCAACGCGCGCTTCCCGCGCAACGCTACTGCATATTCGTTAACCGCT 93
||| |||:|||| |||:|||| |||:|||| |||
Db 257 ValValCysArgSerGlyAlaLeuThrValSerGluIleAlaAlaGlyLeuProAla 276
||| |||:|||| |||:|||| |||:|||| |||
QY 94 TTGTTCTCTCTCAGATAACAAACAAAGTACTGGCAGAGCTTA 144
||| |||:|||| |||:|||| |||:|||| |||
Db 277 IlePheValProPheGlnHisLysAspArgGlnGlnTyrTrpAsnAlaLeu 293
||| |||:|||| |||:|||| |||:|||| |||

Search completed: January 15, 2003, 15:58:36
Job time : 4.88907 secs

Db 876 AspVallysAsnLeuThrGlyLeuValLeuProLysLeuAspArgAspLeuLeuSer 895
 QY 130 TACTGGCAGAGTTAGCAAA 150
 Db 896 SerTrpValAlaLeuAlaGlu 902

RESULT 2
 AD3067
 hypothetical protein Atu4154 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AD3067
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, T.
 ; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClellan
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AD3067
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-317 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AAL44954.1; PID:g17742609; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu4154
 A:Map position: linear chromosome

Alignment Scores:
 Pred. No.: 1.45 Length: 317
 Score: 66.50 Matches: 20
 Percent Similarity: 54.17% Conservative: 6
 Best Local Similarity: 41.67% Mismatches: 15
 Query Match: 24.91% Indels: 7
 DB: 2 Gaps: 3

US-09-856-221-2 (1-154) x AD3067 (1-317)

QY 142 AGCTGTCGCAGTAACCTTTCAGTTTGTATCTGAGGAGGACAAAGCGTTAACG 83
 Db 142 ThrSerAlaHisAsnLeuValAlaLeuGluLysAsnGlyLeuLys---LeuSer 160
 QY 82 AATTGTCAG-----TACGTTTCGCGGCAAGCGCGCGTTCACGAGTTGCGCATTT 29
 Db 161 AspIleAspValThrTrpLeuSerProAla-----AspAlaAlaAlaAlaPhe 176
 QY 28 TATGCTGATAATGATCAGCAATGG 5
 Db 177 AlaSerAspLysIleAspAlaTrp 184

RESULT 3
 F98219
 hypothetical protein AGR_L1399 [imported] - Agrobacterium tumefaciens (strain C58, Cere)
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 C:Accession: F98219
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:11743194
 A:Accession: F98219
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-338 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89280.1; PID:g15159111; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L1399
 A:Map position: linear chromosome

Alignment Scores:

Pred. No.: 1.46 Length: 338
 Score: 66.50 Matches: 20
 Percent Similarity: 54.17% Conservative: 6
 Best Local Similarity: 41.67% Mismatches: 15
 Query Match: 24.91% Indels: 7
 DB: 2 Gaps: 3

US-09-856-221-2 (1-154) x F98219 (1-338)

QY 142 ACGTCTGCCAGTAACCTTTCAGTTTGTATCTGAGGAGGACAAAGCGTTAACG 83
 Db 163 ThrSerAlaHisAsnLeuValAlaLeuGluLysAsnGlyLeuLys---LeuSer 181
 QY 82 AATTGTCAG-----TACGTTTCGCGGCAAGCGCGGTTGACGAGTTGCGCATTT 29
 Db 182 AspIleAspValThrTrpLeuSerProAla-----AspAlaAlaAlaAlaPhe 197
 QY 28 TATGCTGATAATGATCAGCAATGG 5
 Db 198 AlaSerAspLysIleAspAlaTrp 205

RESULT 4

C83104

hypothetical protein PA4327 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: C83104
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: C83104
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-268 <STO>
 A:Cross-references: GB:AE004849; GB:AE004091; NID:g9950550; PIDN:ANG07715.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4327

Alignment Scores:

Pred. No.: 2.31 Length: 268
 Score: 65.00 Matches: 14
 Percent Similarity: 59.45% Conservative: 8
 Best Local Similarity: 37.84% Mismatches: 15
 Query Match: 24.53% Indels: 0
 DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x C83104 (1-268)

QY 13 GATCATTATCAGCATAAAATGCGCACTGCTCAACGCGCGCTTCGCGGCAACGT 72
 Db 132 GluHisAlaGlnAlaArgLeuLeuAlaMetArgGlnAlaArgGluProValcInLys 151
 QY 73 ACTGCATTCGTTACCCCTTTGCTTCCTCCAGATAACAAAACTG 123
 Db 152 SerAlaArgGluLeuAlaThrTrpLysProGlnAsnLeuAspLeu 168

RESULT 5

T07258

cell division protein homolog - Chlorella vulgaris chloroplast
 C:Species: Chloroplast Chlorella vulgaris
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 01-Dec-2000
 C:Accession: T07258; T07259
 R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; S.; Tsudzuki, J.; Na
 proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997
 A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga C
 A:Reference number: Z15985; MUID:97303241; PMID:9159184
 A:Accession: T07258
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-1720 <NA>
A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAAS7905.1; PID:g2224421
A:Accession: T07259
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 791-1720 <NA>
A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAAS7906.1; PID:g2224422
C:Genetics:
A:Genome: chloroplast
A:Note: fish
C:Keywords: chloroplast

Alignment Scores:
Pred. No.: 4.1 Length: 1720
Score: 63.50 Matches: 19
Percent Similarity: 43.40% Conservative: 4
Best Local Similarity: 35.85% Mismatches: 19
Query Match: 23.78% Indels: 11
DB: 2 Gaps: 2

US-09-856-221-2 (1-154) x T07258 (1-1720)

QY 151 GTTGTGCTACGCTGCCAGTAACTTCGAGTTTGTGTTATCTCAGGAGGACAAAG 92
|||||
Db 669 ValIleuAspSerLysLysAsnPhe-----AspTrpSerThrLys 682
QY 91 CGGTTACGAATTTCGAG-----TACGTTTCGCGGCAAGCGCGCGTT 47
|||||
Db 683 HisGlnThrAsnLeuGlnLeuTrpPheGlnLysTrpValSerProLeuAsnProLeuVal 702
QY 46 GACGCGATTGCGCGATTTCATGCTGATAATGATCAGCAA 8
|||||
Db 703 GlnPheGlnGlyAsnPhetPheCysGluGluSerValGlu 715

RESULT 6

AC2293
hypothetical protein all3898 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2293
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Shimpo, S.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; WUID:21595285; PMID:11759840
A:Accession: AC2293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-275 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA875597.1; PID:gl7133032; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all3898

Alignment Scores:
Pred. No.: 5.96 Length: 275
Score: 62.00 Matches: 17
Percent Similarity: 65.00% Conservative: 9
Best Local Similarity: 42.50% Mismatches: 10
Query Match: 23.40% Indels: 4
DB: 2 Gaps: 2

US-09-856-221-2 (1-154) x AC2293 (1-275)

QY 28 AAAATCGGCAACTGCTCAACGCGCGCTTCGCGCGGAA-----CGTACTGCAAT 81
|||||
Db 138 LysMetArgGlnLeuIleLysArgSerProLeuLeuPheLysIleMetArgSerLeuSer 157
QY 82 TCGTTAACCGCTTGTTCCTCTCAGATAACAAAAAAGTCAAGTTACTGCGACG 141
|||||
Db 158 PheLeuThrLysIlePhe-----GlyLeuAspLysLysIleGluGlnAspTrpGlnThr 175

RESULT 7

S77373
hypothetical protein sll1464 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 01-Mar-2002
A:Accession: S77373
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Shimpo, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC 6803
A:Reference number: S74322; WUID:97061201; PMID:8905231
A:Accession: S77373
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-477 <KAN>
A:Cross-references: EMBL:D90906; GB:AB001339; NID:gl652492; PIDN:BAAL7476.1; PID:d101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG
C:Superfamily: hypothetical protein b1706

Alignment Scores:
Pred. No.: 9.86 Length: 477
Score: 60.50 Matches: 19
Percent Similarity: 51.92% Conservative: 8
Best Local Similarity: 36.54% Mismatches: 18
Query Match: 22.83% Indels: 7
DB: 2 Gaps: 3

US-09-856-221-2 (1-154) x S77373 (1-477)

QY 4 ACCATTGCTGATCATATCAGCATAAATGCGCACTG-----CGTCAACGCGGCGC 57
|||||
Db 412 ThrTrpArgSerIleTyGlnGlnValLeuLysGlnLeuProGluSerLeuAlaAla 431
QY 58 TTGCGG---GCGAAACGTTACTGCAATTCGTTAAACGGCTTTCCTCTCAGATAAAC 114
|||||
Db 432 ValProThrValLeuGluGlnAlaAsnProLeuThrAspLeuLeuArgProArgIle--- 450
QY 115 AAAAAATCGCAAGTTACTGCGCAGACGCTTACGACAA 150
|||||
Db 451 -----GluThrValTrpGlnAlaIleAlaGlu 459

RESULT 8

T32189
zinc finger protein nhr-55 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32189
R:Jones, K.; Bradshaw, H.; Graves, T.
Submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T01G6.
A:Reference number: Z21134
A:Accession: T32189
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-455 <JON>
A:Cross-references: EMBL:AF022978; PIDN:AA069068.1; GSPDB:GN00023; CESP:T01G6.7
A:Experimental source: strain Bristol N2; clone T01G6
C:Genetics:
A:Gene: nhr-55; CESP:T01G6.7
A:Map position: 5
A:Introns: 84/2; 128/1; 154/3; 266/3; 333/2; 423/3

Alignment Scores:
Pred. No.: 13.5 Length: 455
Score: 59.50 Matches: 14
Percent Similarity: 55.88% Conservative: 5
Best Local Similarity: 41.18% Mismatches: 8
Query Match: 22.45% Indels: 7
DB: 2 Gaps: 2

```
US-09-856-221-2 (1-154) x T32189 (1-455)
QY 29 AATCGCGCAACGTCGTCAACGCGCGCTGCGCGGGAACGTAAGTCAAAATTCGTTAA 88
Db 85 ArgCysGlySerAlaAsnAsnPhelysProCysArgArgAsnMet----- 99
QY 89 CGGCTTTCCTTCCTCCAGCA--TAAACAAAACACTGCAAA 127
Db 100 ---AsnCysGluPhelLeuLysAsnGlyTrpPheAsnCysLys 112
RESULT 9
F83781
transposase (08) / ABC transporter (ATP-binding protein) BH1054 [imported] - Bacillus ha
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83781
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83781
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-901 <STO>
A:Cross-references: GB:BA000004; NID:g10173440; PIDN:BA04773.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1054
Alignment Scores:
Pred. No.: 14 Length: 901
Score: 59.50 Matches: 13
Percent Similarity: 60.78% Conservative: 18
Best Local Similarity: 25.49% Mismatches: 19
Query Match: 22.45% Indels: 1
DB: 2 Gaps: 1
US-09-856-221-2 (1-154) x F83781 (1-901)
QY 1 CAACACATTCGTATCATTAATGCGCAATGCGCAACGCGGCAACGCGCGCTTG 60
Db 223 GluGlnLeuLysGluThrAlaAspHisLysGluLysGlnAlaGlnLysGlnAlaLeuLys 242
QY 61 CCGCGGAACCTACTGCAAAATTCGTTAACCGCTTGTTCCTCCTCAGATAACAAAAA 120
Db 243 ProLeuArgLysIleTyArgThrLeuThrAspThrLeuProArgLeu---ArgLys 261
QY 121 CTGCAAAATTCGCGACGAGTTAGCACACGC 153
Db 262 TyrGluThrSerPheArgLeuLeuGlyGluArg 272
RESULT 10
AD2325
hypothetical protein all14155 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2325
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075854.1; PID:g17133290; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all14155
US-09-856-221-2 (1-154) x AD2325 (1-278)
QY 7 ATTCGTGATCATTCAGCATATAAATGCGCAACGCGGCTGCGCGGCG 66
Db 170 IleArgAsp-----LeuArgGlnLeuArg---AsnAlaLysIleGluThr 183
QY 67 AAACGTACTCAAAATTCGTTAACGCGCTTGTTCCTTCCT-----CAGATAACAAAAA 120
Db 184 LysIleThrAlaLysSerGluThrGlnGlnGlnIleProArgThrGlnSerAsnIleLys 203
QY 121 CTGCAAAATTCAGTGGCAG 138
Db 204 LeuAspSerCysTrpGln 209
RESULT 11
B71361
Probable thiamin ABC transporter, thiamin-binding periplasmic protein - syphilis spir
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 03-Jun-2002
C:Accession: B71361
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: B71361
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-335 <COL>
A:Cross-references: GB:AE001199; GB:AE000520; NID:g3322402; PIDN:AAC65133.1; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0144
C:Superfamily: sfua protein
Alignment Scores:
Pred. No.: 15.5 Length: 335
Score: 59.00 Matches: 16
Percent Similarity: 52.63% Conservative: 4
Best Local Similarity: 42.11% Mismatches: 18
Query Match: 22.10% Indels: 0
DB: 2 Gaps: 0
US-09-856-221-2 (1-154) x B71361 (1-335)
QY 136 GCCAGTAACTTTCAGCTTTTGTATTCAGGAGGAACAAAGCGTTAACGAATTG 77
Db 271 AlaThrGlnPhleAspPheLeuLeuSerThrGluAlaGlnGluLeuProLeuThr 290
QY 76 CAGTACGTTTCGCGGCAAGCGCGCTTCACCCACTTGCCTTCCTTATGCT 23
Db 291 GlnTrpMetTyrProAlaAsnProAspValProLeuProAlaAlaPheSerAla 308
RESULT 12
CB87664
methyl-accepting chemotaxis protein McpH [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 01-Mar-2002
C:Accession: CB87664
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
```

```

A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87664
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-589 <STO>
A:Cross-references: GB:AE005673; NID:gl3425051; PIDN:AAK25311.1; GSPDB:GN00148
C:Genetics:
C:Superfamily: methyl-accepting chemotaxis protein mcpA

Alignment Scores:
Pred. No.: 18.8 Length: 589
Score: 58.50 Matches: 17
Percent Similarity: 52.17% Conservative: 7
Best Local Similarity: 36.96% Mismatches: 17
Query Match: 21.91% Indels: 5
DB: 2 Gaps: 2

US-09-856-221-2 (1-154) x C87664 (1-589)
QY 154 GCCTGTGCTAACGCTCCAGTACCTTCAGTT-----TTTGTGTTATCTGAG 104
DB 183 GlyValIleLeuAsnAlaAsnGluAsnPhelEuAlaThrValGlyTyrAlaLeuSerGlu 202
QY 103 GAAGGACAAAGCGGTTCAGAAATTCGAGTACGTTCCCGCGAAGCGCGCGTTGAC 44
DB 203 IleGlnGlyLysHis-----HisSerMetPheValAspProAlaGluAlaArgSerGlu 220
QY 43 GCAGTTCGCCGCAATTTAT 26
DB 221 AlaTyrAlaAlaPheTrp 226

RESULT 13
F81896
probable phage transposase NNA1284 [imported] - Neisseria meningitidis (strain Z2491 ser
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: F81896
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: F81896
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-657 <PAR>
A:Cross-references: GB:AL157959; GB:AL162755; NID:g7379742; PIDN:CAB84536.1; PID:g737996
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NNA1284

Alignment Scores:
Pred. No.: 18.9 Length: 657
Score: 58.50 Matches: 14
Percent Similarity: 47.83% Conservative: 8
Best Local Similarity: 30.43% Mismatches: 21
Query Match: 22.08% Indels: 3
DB: 2 Gaps: 1

US-09-856-221-2 (1-154) x F81896 (1-657)
QY 16 CATTATACAG-----CATAAATGCGCAACTGCGTCAACGCGCGCGCTCCGCGG 66
DB 485 HistYrSerProLysAlaTyrArgGluMetArgLeuGluGlnAspGlyIleAlaProasp 504
QY 67 AAACGTACTGCAATTCGTTACCGCTTTGTTCCCTCAGCATAAACAAAACACTCAA 126
DB 505 MetLeuSerAlaGluLeuAlaThrMetPheMetProGlnGluValArgLysValGln 524
QY 127 AGTTACTGGCAGACGTA 144
DB 525 ArgGlyTyrLeuAspLeu 530

RESULT 14
B82083
conserved hypothetical protein VC2387 [imported] - Vibrio cholerae (strain N16961 ser
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82083
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
J., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82083
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <HEI>
A:Cross-references: GB:AE004309; GB:AE003852; NID:g9656954; PIDN:AAF95530.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2387
A:Map position: 1

Alignment Scores:
Pred. No.: 20.8 Length: 205
Score: 58.00 Matches: 14
Percent Similarity: 47.83% Conservative: 8
Best Local Similarity: 30.43% Mismatches: 16
Query Match: 21.72% Indels: 8
DB: 2 Gaps: 2

US-09-856-221-2 (1-154) x B82083 (1-205)
QY 133 ACTAATCTTCCAGTTTTCGTTATCTCAGCAA-----CGAACAAAC 92
DB 87 SerAsnTyrAlaValSerCysIleGlnGluAsnArgLysThrGlnThrCysGlyGlnLys 106
QY 91 CGGTTACGAATTCAGTACGTTTCGCGGCAAGCGCGCGTTCGACGCGGTCGCCGA 32
DB 107 LysLysAspAsnIleLeuPheIleGluProThrSerProSerCysAsn-----SerHis 124
QY 31 TTTTATGCTGATAATGAT 14
DB 125 PheSerLeuAspThrAsp 130

RESULT 15
G71006
probable thiamin-binding periplasmic protein precursor - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: G71006
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71006
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-347 <KAN>
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30455.1; PID:g3237772
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH1349
C:Superfamily: sfuA protein

Alignment Scores:
Pred. No.: 21.4 Length: 347
Score: 58.00 Matches: 16
Percent Similarity: 40.32% Conservative: 9
Best Local Similarity: 25.81% Mismatches: 19
Query Match: 21.72% Indels: 18

```

```

DB:          2          Gaps:          2
US-09-856-221-2 (1-154) x G71006 (1-347)
QY 136 GCCAGTAACTTTCAGTCTTTTTCAGTCTGAGGAGCAAGCGGTTAACGAATTG 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 AlaGluArgPheValGluPheLeuIleSerGlnLysAlaGlnGluLysLeuProThrThr 287
QY 76 CAGTACGTTTCGCCGGCA-----AGGCCG 53
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 GlnTrpMetPheProAlaAsnLysValLysLeuProGluValTyLysTyAsnPro 307
QY 52 CGCGTTCAGCAGTTCGCCGATTTTATGCTGAT-----AATGATCAC 11
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 LysValAspPheSerLysAlaValTyLeuAspProLysLeuIleGlnGluAsnTyGlu 327
QY 10 GAATGG 5
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 LysTrp 329

```

Search completed: January 15, 2003, 15:56:19
Job time : 11.0971 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 3.64274 Seconds
(without alignments)
3506.895 Million cell updates/sec

Title: US-09-856-221-2

Perfect score: 265
Sequence: 1 caaacattcgtatcatta.....gcagacgttagcacacgcc 154

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p_model -DEV=xlsp
-O=/cnp2.1/USP70.spool/US09856221/runat_15012003.153924.12445/app_query.fasta.1.1948
-DB=SpisProt_40 -OPMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPT=0 -LOOPT=0
-UNIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221.ecgn.1.138.runat_15012003.153924.12445 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	25.7	1718	1 RRPO_SHVX	Q04575 shallot vir
2	63.5	23.8	1720	1 FTSH_CHLVU	P56369 chlorella v
3	60.5	22.8	477	1 Y264_SYNY3	P73436 synchocyst
4	59.5	22.5	455	1 NH55_CAEEL	O16962 caenorhabdi
5	58.5	21.9	490	1 MET3_SCHPO	P78937 schizosacch
6	58	21.9	694	1 LCF3_YEAST	P39002 saccharomyc
7	58	21.9	767	1 HTPF_SYNY3	Q56638 synchocyst
8	57.5	21.5	87	1 Y07A_BPT4	P39226 bacterioph
9	57	21.3	322	1 MYDM_HUMAN	Q07494 gallus gall
10	56.5	21.2	984	1 EPB1_CHICK	Q07494 gallus gall
11	56	21.1	354	1 MURG_ECOLI	Q8x9Y8 escherichia
12	56	21.1	354	1 YFIQ_ECOLI	P76594 escherichia
13	56	21.1	886	1 CATG_MOUSE	P28293 mus musculu
14	55.5	20.9	261	1 VP3_ROTPE	P36192 porcine rot
15	55.5	20.9	692	1 Y290_LAMB2	P03766 bacterioph
16	55	20.6	290	1 Y290_LAMB2	P49907 bos taurus
17	55	20.8	402	1 SELP_BOVIN	O74657 candida alb
18	55	20.8	468	1 ALS2_CANAL	

19	55	20.8	828	1 YKR6_YEAST	P34239 saccharomyc
20	55	20.8	888	1 SM6A_MOUSE	O35464 mus musculu
21	55	20.8	1260	1 ALS1_CANAL	P46590 candida alb
22	55	20.8	1419	1 ALAI_CANAL	O13368 candida alb
23	54	20.2	201	1 YAH2_ECOLI	P77736 escherichia
24	54	20.4	294	1 AMPM_METUA	O58725 methanococc
25	54	20.2	296	1 MYDM_MOUSE	O35682 mus musculu
26	54	20.4	540	1 Z136_HUMAN	P52737 homo sapien
27	54	20.2	2733	1 RRPB_CVMA5	P16342 murine coro
28	53.5	20.2	187	1 UCRI_RHOSH	Q02762 rhodobacter
29	53.5	20.0	270	1 CB12_PETHY	P13869 petunia hyb
30	53.5	20.0	340	1 RIR2_MYCGE	P47471 mycoplasma
31	53.5	20.0	352	1 RADA_METUA	O49593 methanococc
32	53.5	20.2	357	1 CARA_LALIC	Q914N5 lactococcu
33	53.5	20.0	420	1 MP13_MESAU	P48968 mesocricetu
34	53.5	20.2	424	1 GCH2_CHLTR	O84736 chlamydia t
35	53.5	20.0	623	1 EXAA_PSEAE	Q924J7 pseudomonas
36	53	19.9	312	1 YPRL_YEAST	O12458 saccharomyc
37	53	19.9	330	1 FBP_NEIGO	P17259 neisseria g
38	53	20.0	354	1 MURG_SALTI	Q8Z9G9 salmonella
39	53	20.0	354	1 MURG_SALTI	Q8ZRU3 salmonella
40	53	20.0	1119	1 ALS3_CANAL	O74623 candida alb
41	53	19.9	2731	1 RRPB_CVMAJH	P29982 murine coro
42	52.5	19.8	266	1 TERM_BPO3	O37883 bacterioph
43	52.5	19.8	327	1 COL7_ARATH	Q9SYM2 arabidopsis
44	52.5	19.7	339	1 RIR2_MYCPN	P75461 mycoplasma
45	52.5	19.7	354	1 PONI_MOUSE	P52430 mus musculu

ALIGNMENTS

RESULT 1

ID	RRPO_SHVX	STANDARD	PRT	1718 AA.
AC	O04575			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	RNA replication protein (194 kDa protein) (ORF 1) [Contains: RNA-			
DE	directed RNA polymerase (EC 2.7.7.48); Probable helicase].			
OS	Shallot virus X (ShVX).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Allexivirus.			
NCBI_TaxID=31770;				
[1]				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93019008; PubMed=1339468;			
RA	Karyuka K.V., Vishnichenko V.K., Levay K.E., Kondrikov D.Y.,			
RA	Ryabov E.V., Zavyetov S.K.			
RT	"Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal			
RT	cistron closely related to those of potexviruses and a unique			
RT	arrangement of the 3'-proximal cistrons."			
RL	J. Gen. Virol. 73:2553-2560(1992).			
CC	-!- FUNCTION: RNA-replication. The central part of this protein			
CC	possibly functions as an ATP-binding helicase.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +			
CC	{RNA}(N).			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M97264; AAA47787.1; -
PIR; J01734; J01734.
InterPro; IPR005123; 2OG-feil_Oxy.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR006060; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF03171; 2OG-feil_Oxy; 1.
DR

DR Pfam: PF01434; Peptidase_M41; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00674; AAA; 1.
KW Cell division; Atp-binding; Transmembrane; Chloroplast.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 896 916 POTENTIAL.
FT TRANSMEM 973 993 POTENTIAL.
FT TRANSMEM 1021 1041 POTENTIAL.
SQ SEQUENCE 1720 AA; 197172 MW; OFLEA926B799D5BB CRC64;

Alignment Scores:
Pred. No.: 176 Length: 1720
Score: 63.50 Matches: 19
Percent Similarity: 43.40% Conservative: 4
Best Local Similarity: 35.85% Mismatches: 19
Query Match: 23.78% Indels: 11
DB: 1 Gaps: 2

US-09-856-221-2 (1-154) x FTSH_CHLVU (1-1720)

QY 151 GTTCGTGCAACGCTGCCAGTAACCTTTGCAGTTTTTGTATTCTGAGGAAGAACAAAG 92
|||:::||| ||| |||||| |::: |||||
Db 669 VallileuAspSerIysAsnPh-----AspTpSerThrIys 682
QY 91 CGTTTACGAATTTGCAG-----TACGTTTCGGCGGCAAGCCGCGCCTT 47
|||||:||||| ||||| ||||| |||||
Db 683 HisGlnTrasnLglnLeuTrpPheGlnLysTyValSerProLeuAsnProLeuVal 702
QY 46 GACGCAGTTGCCCATTTTATGCTGATAATGATCACGAA 8
|||::: |||
Db 703 GlnPheGlnGlyAsnPhPheCysGluGluSerValGlu 715
RESULT 3
Y264.SYN3 STANDARD; PRT; 477 AA.
ID Y264.SYN3
AC P73436;
DT 15-JUL-1998 (Rel. 36, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein sll1464.
SL SLL1464.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugihara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Maruta A., Nakazaki N., Naito K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D90906; BAAL17476.1; -.
CC InterPro: IPR003846; UPF0061.
CC Param: PF02696; UPF0061; 1.
CC Hypothetical protein; Complete proteome.
SK SQ SEQUENCE 477 AA; 54041 MW; 81F6899B1A6D613C CRC64;

Alignment Scores:


```

Pred. No.: 4.18 Length: 477
Score: 60.50 Matches: 19
Percent Similarity: 51.92% Conservative: 8
Best Local Similarity: 36.54% Mismatches: 18
Query Match: 22.83% Indels: 7
DB: 1 Gaps: 3

US-09-856-221-2 (1-154) x YZ64_SYN3 (1-477)
QY 4 ACCATTCGTGATCATTAATGCGCAACTG-----CGTCAACGGCGGCC 57
DB 412 thrrpgrserleryglnglnvaldeuylsglnleuprrogluSerLeuAlaLa 431
QY 58 TTGCGG---CGGAACACTACTGCAATTCGTTAACCGCTTTCTTCCTCAGATAAAC 114
DB 432 ValProThrValLeuGluGlnAlaAsnProLeuThrAspLeuLeuArgProArgile--- 450
QY 115 AAAAAAAGTCTACTGCGCAGACGTTAGCACAA 150
DB 451 -----GluThrValTrpGlnAlaIleAlaGlu 459

RESULT 4
NH55_CAEEL STANDARD; PRT; 455 AA.
ID NH55_CAEEL
AC Q16962: Q9GTF7; Q9GTF8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Nuclear hormone receptor family member nhr-55.
GN NHR-55 OR T01G6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Jones K., Bradshaw H., Graves T.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 103-455 FROM N.A.
RA Bogan A., Malina C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
compatibility with the ligand-binding domain fold."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF022978; AAC24179.1;
CC EMBL: AF273801; AAG15150.1;
CC EMBL: AF273802; AAG15151.1;
CC HSPF: P20393; 1A6Y.
CC WormPep: T01G6.7; CEI3002.
CC InterPro: IPR000536; Hormone_rec_lig.
CC InterPro: IPR001628; znf_C4steroid.
CC Pfam: PF00105; znf_C4; 2.
CC PRINTS: PR000047; STROIDFINGER.
CC ProDom: PD000035; znf_C4steroid; 1.
CC SMART: SM00430; HOLI; 1.
CC SMART: SM00359; znf_C4; 1.
CC PROSITE: PS00031; NUCLEAR_RECEPTOR; FALSE_NEG.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
DB DNA_BIND 59 124 NUCLEAR RECEPTOR-TYPE.

```

```

FT ZN_FING 59 86 C4-TYPE.
FT ZN_FING 95 119 C4-TYPE.
FT DOMAIN 34 40 POLY-SER.
FT DOMAIN 43 50 POLY-SER.
SQ SEQUENCE 455 AA; 52956 MW; B56CEP5A00D6B1A3 CRC64;

Alignment Scores:
Pred. No.: 5.71 Length: 455
Score: 59.50 Matches: 14
Percent Similarity: 55.88% Conservative: 5
Best Local Similarity: 41.18% Mismatches: 8
Query Match: 22.45% Indels: 7
DB: 1 Gaps: 2

US-09-856-221-2 (1-154) x NH55_CAEEL (1-455)
QY 29 AAATGGCGCAACTGCTCAACGGCGGCTTCGCCGGAACGTAAGTGTGCAAAATTCGTTAA 88
DB 85 ArgCysGlySerAlaAsnAsnPhelLysProCysArgArgAsnMet----- 99
QY 89 CCGCTTTGTTCTCTCTCAGCA---TAAACAAAACAACTGCAAA 127
DB 100 ---AsnCysGluPheLeuLysAsnGlyTrpPheAsnCysLys 112

RESULT 5
MET3_SCHPO STANDARD; PRT; 490 AA.
ID MET3_SCHPO
AC P78937; Q9USV5; Q9P6S1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sulfate adenylyltransferase (EC 2.7.7.4) (Sulfate adenylylate
transferase) (SAT) (ATP-sulfurylase).
GN SPC28F2.01C OR SPC27.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Kohnosu A., Niwa O., Yano M., Saitoh S., Katayama T., Nagao K.,
RA Yanagida M.;
RT "S pombe chromosome II cosmid 1228 sequence."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy I., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.A., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

```


RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDIJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (probable).
 CC -!- SIMILARITY: BELONGS TO THE MAL FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AY037147; AAK67628.2; -;
 DR EMBL; BC013995; AAH13995.1; -;
 KW Transmembrane.
 FT TRANSMEM 41 61 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 171 191 POTENTIAL.
 FT TRANSMEM 203 223 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 SQ SEQUENCE 322 AA; 35273 MW; A92142CC63625A4E CRC64;
 Alignment Scores:
 Pred. No.: 12.3 Length: 322
 Score: 57.00 Matches: 15
 Percent Similarity: 55.00% Conservative: 7
 Best Local Similarity: 37.50% Mismatches: 18
 Query Match: 21.35% Indels: 0
 DB: 1 Gaps: 0
 US-09-856-221-2 (1-154) x MYDM_HUMAN (1-322)
 QY 145 CTAAGCTCGCAGTACCTTCAGTCTTTTGTATCTGAGGAGCAACGCGTTA 86
 DB 103 IletHrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyrProThr 122
 QY 85 ACGAAATTCAGTACGTTTGGCGGCAAGCGCGGTTGACGAGTTCGCGCATTTTAT 26
 DB 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgHisAlaIleAlaAlaThrPhe 142
 RESULT 10
 EPBL_CHICK
 ID EPBL_CHICK STANDARD; PRT; 984 AA.
 AC Q07494;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ephrin type-B receptor 1 (EC 2.7.1.112) (Tyrosine-protein kinase
 DE receptor Eph-2) (tyrosine kinase CER6 receptor) (fragment).
 GN EPHB1 OR CER6.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus
 OC NCBI_TaxID=9031;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93288394; PubMed=8510926;
 RA Sajjadi F.G., Pasquale E.B.;
 RT "Five novel avian Eph-related tyrosine kinases are differentially
 RT expressed."
 RL Oncogene 8:1807-1813(1993).
 CC -!- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-B FAMILY. BINDS TO
 CC EPHRIN-B1, B2 AND B3. MAY BE INVOLVED IN CELL-CELL INTERACTIONS
 CC IN THE NERVOUS SYSTEM.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE 10-DAY EMBRYO,
 CC AND IN ADULT BRAIN, LUNG, HEART AND SKELETAL MUSCLE. LOW LEVELS OF
 CC EXPRESSION DETECTED IN ALL OTHER ADULT TISSUES TESTED.
 CC -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN
 CC RECEPTOR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z19110; CAA79526.1; -;
 DR HSSP; P29323; IBAF.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001090; Ephrin_receptor.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003962; FNIII_repeat.
 DR InterPro; IPR001660; SAM.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR001426; Ykase_receptorV.
 DR Pfam; PF00041; fn3_2.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00536; SAM; 1.
 DR Pfam; PF01404; EPH_lbd; 1.
 DR PRINTS; PR00014; FNTYPEIII.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR ProDom; PD001495; Ephrin_receptor; 1.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00454; SAM; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00791; RECEPTOR_TYR_KIN_V_1; PARTIAL.
 DR PROSITE; PS00791; RECEPTOR_TYR_KIN_V_2; 1.
 DR PROSITE; PS0105; SAM_DOMAIN; 1.
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Receptor; Transmembrane; Glycoprotein; Repeat.
 FT NON_TER 1 1
 FT DOMAIN <1 541 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 542 562 POTENTIAL.
 FT DOMAIN 563 984 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 164 300 CYS-RICH.
 FT DOMAIN 301 410 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 411 527 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 619 882 PROTEIN KINASE.
 FT DOMAIN 911 975 SAM.
 FT SITE 982 984 PDZ-BINDING MOTIF (POTENTIAL).
 FT NP_BIND 625 633 ATP (BY SIMILARITY).
 FT BINDING 651 651 ATP (BY SIMILARITY).
 FT ACT_SITE 744 744 BY SIMILARITY.
 FT MOD_RES 594 594 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT MOD_RES 600 600 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT MOD_RES 778 778 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT MOD_RES 928 928 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 SQ SEQUENCE 984 AA; 109519 MW; EF06C83B863A13A1 CRC64;
 Alignment Scores:
 Pred. No.: 15.4 Length: 984
 Score: 56.50 Matches: 17
 Percent Similarity: 41.51% Conservative: 5
 Best Local Similarity: 32.08% Mismatches: 27
 Query Match: 21.16% Indels: 4
 DB: 1 Gaps: 1

```

US-09-856-221-2 (1-154) x EPBL_CHICK (1-984)
QY 152 CGTTGCTAACGCTGCGCAGTAACCTTTCAGGTTTTT-----TGTATTATCTGAGGA 102
Db 176 LysCysProSerValValGln-AsnPheAlaIlePheProGluThrMetThrGlyAlaG1 195
QY 101 AGGAACAACGCGTTTAAAGCAATTTGCGAGTTCGCGCGCAAGCGCGCTTCACGC 42
Db 195 userThrSerLeuValThrAlaAaGlyThrCysIleProAsnAlaGluGluValAspVa 215
QY 41 AGTTGCCGCGATTTTATGCTGATATGATACACGAATGG 5
Db 215 lProIleLysLeuTyrcysAsnGlyAspGlyGluTrp 227

RESULT 11
MURG_ECO57
ID MURG_ECO57 STANDARD: PRT; 354 AA.
AC Q8X9Y8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-) (Undecaprenyl-pp-MurNac-pentapeptide-UDPglcNac glcNac
DE transferase).
GN MURG OR Z0100 OR ECS0094.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=83334;
RX STRAIN=O157:H7 / ED1933 / ATCC 700927;
MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Iii, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;
RA Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tooe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA *Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.*;
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNac (lipid intermediate II) (By similarity).
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF005185; AAC54394.1;
CC EMBL: AP002550; BAB33517.1;
CC InterPro: IPR004276; Glyco_transf_28.

DR Pfam: PF03033; Glyco_transf_28; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 354 AA; 37669 MW; 8C4617F0BBA10DE9 CRC64;

Alignment scores:
Pred. No.: 16.9 Length: 354
Score: 56.00 Matches: 18
Percent Similarity: 37.14% Conservative: 8
Best Local Similarity: 25.71% Mismatches: 22
Query Match: 21.13% Indels: 22
Db: 1 Gaps: 2

US-09-856-221-2 (1-154) x MURG_ECO57 (1-354)
QY 1 CAACACATTCGTGATCATTTAT-----CAGCATAAATGGCGCAACTG 42
Db 223 GlnSerValGluGlnAlaTyraAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 242
QY 43 -----CGTCAACGCGC 54
Db 243 IleAspAspMetAlaAlaAlaTyraAlaTyraAlaAspValValValCysArgSerGlyAla 262
QY 55 GCCTTCCGCGCAAGCTACTGCAATTCGTTAACCCTTGTTCCTCTCTCAGTAAAC 114
Db 263 LeuThrValSerGluIleAlaAlaAlaGlyLeuProAlaLeuPheValProPheGlnHis 282
QY 115 AAAAACTGCAAGTTACTGCGACAGCTTA 144
Db 283 LysAspArgGlnGlnTyTrpAsnAlaLeu 292

RESULT 12
MURG_ECOLI
ID MURG_ECOLI STANDARD: PRT; 354 AA.
AC P17443;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1992 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(Pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-) (Undecaprenyl-pp-MurNac-pentapeptide-UDPglcNac glcNac
DE transferase).
GN MURG OR B0090.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90326550; PubMed=2197603;
RA Ikeda M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;
RA *Nucleotide sequence involving murG and murC in the mra gene cluster
RT region of Escherichia coli.*;
RL Nucleic Acids Res. 18:4014-4014(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RA *Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.*;
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

```

RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE OF 8-354 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90251461; PubMed=2187180;
 RA Mengin-Lecreulx D., Texier L., van Heijenoort J.;
 RT "Nucleotide sequence of the cell-envelope murg gene of *Escherichia coli*.";
 RL Nucleic Acids Res. 18:2810-2810(1990).
 RN [5]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=93194811; PubMed=8449890;
 RA Bupp K., van Heijenoort J.;
 RT "The final step of peptidoglycan subunit assembly in *Escherichia coli* occurs in the cytoplasm.";
 RL J. Bacteriol. 175:1841-1843(1993).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=20348887; PubMed=10892798;
 RA Ha S., Walker D., Shi Y., Walker S.;
 RT "The 1.9 A crystal structure of *Escherichia coli* Murg, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis.";
 RL Protein Sci. 9:1045-1052(2000).
 RN [1]
 RP FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAc-PENTAPEPTIDE (LIPID INTERMEDIATE 1) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAc-PENTAPEPTIDE GLCNAC (LIPID INTERMEDIATE 11).
 CC -1- PATHWAY: Peptidoglycan biosynthesis; last step.
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.
 CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X52644; CAA36867.1; -.
 DR EMBL; X52540; CAA36776.1; -.
 DR EMBL; X55034; CAA38867.1; -.
 DR EMBL; D10483; BAA01355.1; -.
 DR EMBL; AE000118; AAC73201.1; -.
 DR PIR; JH0093; BVRCMG.
 DR PIR; S40600; S40600.
 DR PDB; 1F0K; 27-JUL-00.
 DR EcoGene; EG10623; murg.
 DR InterPro; IPR004276; Glyco_transf_28.
 DR Pfam; PF03033; Glyco_transf_28; 1.
 DR TIGRFAMs; TIGR01133; murg; 1.
 KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
 KW Inner membrane; Peptidoglycan synthesis; Complete proteome;
 KW 3D-structure.
 FT INIT MET 0 0
 SQ SEQUENCE 354 AA; 37683 MW; 0E3FAD945D769C1D CRC64;

Query Match: 21.13% Indels: 22
 DB: 1 Gaps: 2
 US-09-856-221-2 (1-154) x MURG_ECOLI (1-354)
 QY 1 CAACACATTCGTGATCAATAT-----CAGCATAAATCGCGCAACTG 42
 DB 223 GlnSerValGluGlnAlaValAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 242
 QY 43 -----CGTCAACGCGG 54
 DB 243 IleAspAspMetAlaAlaAlaValAlaValAlaValAlaValAlaValAlaValAla 262
 QY 55 GCCTTCGCGCGGAAAGTACGTAATTCGTAACCGCTTGTCTCTCCTCAGATAAAC 114
 DB 263 LeuThrValSerGluLeuAlaAlaAlaGlyLeuProAlaLeuPheValProPheGlnHis 282
 QY 115 AAAAACTGCAGAGTTACTTGGCAGACGTTA 144
 DB 283 LysAspArgGlnGlnTyrTrpAsnAlaLeu 292
 RESULT 13
 YFIQ_ECOLI STANDARD; PRT; 886 AA.
 ID YFIQ_ECOLI
 AC P76594; Q47320;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Hypothetical protein yfiQ.
 GN YFIQ OR B2584.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MGI1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE OF 1-612 FROM N.A.
 RC STRAIN=K12;
 RA Nashimoto H., Saito N.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME, IN THE N-TERMINAL, TO SUCCINYL-COA LIGASES.
 CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ0590.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE000344; AAC75637.1; -.
 DR EMBL; D64044; BAA10925.1; -.
 DR EcoGene; EG14224; yfiQ.
 DR InterPro; IPR003781; COA_binding.
 DR InterPro; IPR000182; GCN5acetyltransf.
 DR Pfam; PF00583; Acetyltransf; 1.
 DR Pfam; PF02629; CoA_binding; 1.
 DR Hypothetical protein; Complete proteome.
 FT CONFLICT 506 507 ST -> YA (IN REF. 2)
 SQ SEQUENCE 886 AA; 97987 MW; 2559449E2961251 CRC64;

Alignment Scores:
 Pred. No.: 17.9 Length: 886

Score: 16.9 Length: 354
 Percent Similarity: 56.00 Matches: 18
 Best Local Similarity: 37.14% Conservative: 8
 Mismatches: 22

```
Score: 56.00 Matches: 12
Percent Similarity: 60.00% Conservative: 6
Best Local Similarity: 40.00% Mismatches: 12
Query Match: 21.13% Indels: 0
DB: 1 Gaps: 0

US-09-856-221-2 (1-154) x YFIQ_ECOLI (1-886)
QY 19 TATCAGATAAAATGGCGCAACTGCTCAACGCGCGCTTGGCGGCAACGTAAGTCTGCA 78
    |||...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||...:
Db 442 TyrAgaGAsnglnGlnLeuAgluThrProAlaLeuProSerAsnLeuThrSer 461
QY 79 AATTCGTTAAACCGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 108
    |||...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||...:
Db 462 AsnThrAlaGluAlaHisLeuLeuLeuGln 471

RESULT 14
CATG_MOUSE
ID CATG_MOUSE STANDARD; PRT; 261 AA.
AC P28293;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin G precursor (EC 3.4.21.20) (Vimentin-specific protease)
DE (VSP).
DE CTSG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Connective tissue;
RX MEDLINE=93200524; PubMed=8453108;
RA Heusel J.W., Scarpati E.M., Jenkins N.A., Gilbert D.J., Copeland N.G.,
RA Shapiro S.D., Ley T.J.;
RT "Molecular cloning, chromosomal location, and tissue-specific
RT expression of the murine cathepsin G gene.";
RL Blood 81:1614-1623(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Kulmburg P., Baumruker T., Werner F.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Huang R., Aveskogh M., Hellman L.T.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 21-60.
RX MEDLINE=92249339; PubMed=1577012;
RA Nakamura N., Tsuru A., Hirayoshi K., Nagata K.;
RT "Purification and characterization of a vimentin-specific protease in
RT mouse myeloid leukemia cells. Regulation during differentiation and
RT identity with cathepsin G.";
RL Eur. J. Biochem. 205:947-954(1992).
CC -1- FUNCTION: THIS VIMENTIN-SPECIFIC PROTEASE MAY REGULATE THE
CC REORGANIZATION OF VIMENTIN FILAMENTS, OCCURRING DURING CELL
CC DIFFERENTIATION, MOVEMENT AND MITOSIS
CC -1- CATALYTIC ACTIVITY: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.
CC -1- SUBCELLULAR LOCATION: STRONGLY ASSOCIATED WITH MEMBRANES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; W96801; AAA37376.1; -
DB EMBL; X70057; CAA49661.1; -
```

```
DR EMBL; X78544; CAA55290.1; -.
DR PIR; S23170; S23170.
DR HSSP; P08311; ICGH.
DR MEROPS; S01.133; -.
DR MGD; MGI:88563; Ctsq.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRY_PSP; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal;
KW Intermediate filament; Membrane.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 20 ACTIVATION PEPTIDE.
FT CHAIN 21 261 CATHEPSIN G.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 142 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT CARBOHYD 71 71 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 51 51 G->S (IN REF. 4).
FT CONFLICT 56 56 E->G (IN REF. 4).
FT CONFLICT 60 60 L->P (IN REF. 4).
SQ SEQUENCE 261 AA; 29095 MW; 5EFA1A6E10E1D7FC CRC64;

Alignment Scores:
Pred. No.: 19.4 Length: 261
Score: 55.50 Matches: 17
Percent Similarity: 57.14% Conservative: 7
Best Local Similarity: 40.48% Mismatches: 11
Query Match: 20.94% Indels: 7
DB: 1 Gaps: 2

US-09-856-221-2 (1-154) x CATG_MOUSE (1-261)
QY 1 CAACCAATCGTGATCATATATCAGCATATAAATCGGCAACTCGTCAACGCGCGCTTG 60
    ||| |||||...: ||| |||||...: ||| |||||...: |||||
Db 103 GlnAsnIleArgAsnAspIleMet-----LeuLeuGlnLeuArgArgAla----- 118
QY 61 CGGCGAAACGTACTGCAATTCGTTAACCCTTGTCTCTCTCTCTCTCTCTCTCTCTCT 120
    ||| |||...: |||...: |||...: |||...: |||...: |||...: |||...:
Db 119 -----ArgArgSerGlySerValIysProValAlaLeuProGlnAlaSerLysLys 135
QY 121 CTGCAA 126
    |||||
Db 136 LeuGln 137

RESULT 15
VP3_ROTPC
ID VP3_ROTPC STANDARD; PRT; 692 AA.
AC P26192;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Inner core protein VP3.
GN S4.
OS Porcine rotavirus (group C / strain Cowden).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10916;
RN [1]
RX MEDLINE=92124743; PubMed=1310192;
RA Bremont M., Juste-Lesage P., Chabanne-Vautherot D.,
RA Charpillienne A., Cohen J.;
RT "Sequences of the four larger proteins of a porcine group C rotavirus
RT and comparison with the equivalent group A rotavirus proteins.";
RL Virology 186:684-692(1992).
CC -1- FUNCTION: BINDS GTP AND COULD PLAY A ROLE IN THE CAPPING
```

CC MECHANISM (BY SIMILARITY).
CC -1- SURCELLULAR LOCATION: INNER CORE.
CC -1- SIMILARITY: TO OTHER VP3 PROTEINS.
CC -1- SIMILARITY: SIGNIFICANT, TO THE PROKARYOTIC ADENINE-SPECIFIC
CC METHYLTRANSFERASE FORK PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M74219; AAA99239.1; -.
DR PIR; C40822; P3XRPC.
KW GTP-binding.
SQ SEQUENCE 692 AA; 81378 MW; D679352CF22FE7DE CRC64;

Alignment Scores:
Pred. No.: 20.6 Length: 692
Score: 55.50 Matches: 16
Percent Similarity: 43.64% Conservative: 8
Best Local Similarity: 29.09% Mismatches: 16
Query Match: 20.94% Indels: 15
DB: 1 Gaps: 2

US-09-856-221-2 (1-154) x VP3_ROTTC (1-692)
Qy 16 CATTATCAGCAATAAATCGCGCAACTGCGTCAACGCGCGCTTGGCGGCAACGTACT 75
Db 183 AsnPhenAlaHisPheMetArgLeuLeuArgMetArgPheAlaValPropPheAspGlnLeu 202
Qy 76 GCAAAATCGTTAAC-----GCTTTGTC----- 99
Db 203 SerAsnArgValThrArgSerArgAlaPhePheLysSerLysIleHisIleGlyLeuArg 222
Qy 100 -----CTTCTCAGATAACAAACTGCAAAAGTACTGG 135
Db 223 AsnGluSerIleProGlnAlaLeuAspAsnIleAsnSerGlnTrp 237

Search completed: January 15, 2003, 15:46:27
Job time : 6.64274 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:44:13 : Search time 14.5082 Seconds
(without alignments)
4374.264 Million cell updates/sec

Title: US-09-856-221-2

Perfect score: 265

Sequence: 1 caaacattcgatcatta.....gcagacgttagcacacgcc 154

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=klp
-O=/gen2_1/USPTO_SPOOL/US09856221/runat_15012003_153925_12457/app_query.fasta_1.1948
-DB=SPTRMBL_21_QWMT-us09856221-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -XLEN=2000000000
-USER=US09856221@cgn_1_171-runat_15012003_153925_12457 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGESUBQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	233	87.9	2523	2	Q93RP0	Q93rp0 xenorhabdus

2	137	51.7	2376	2	Q9F9Z3
3	126.5	47.7	2504	2	Q9F9Z3
4	118.5	44.7	2516	2	Q9RN43
5	102	38.5	2538	2	Q93RN7
6	67	25.3	1625	12	Q95597
7	66.5	24.9	338	16	Q8U8D8
8	65	24.5	268	16	Q9HW74
9	63.5	24.0	886	5	Q9VGC3
10	63	23.8	1550	12	Q95601
11	62	23.4	275	16	Q8YOD5
12	62	23.2	339	2	Q9R2H9
13	61	22.8	108	12	Q8DH37
14	61	23.0	1557	12	Q8QXW5
15	60.5	22.8	289	3	Q9F696
16	60.5	22.8	658	5	Q9NH88
17	60.5	22.8	658	5	Q9NFR7
18	60.5	22.8	658	5	Q9VB22
19	59.5	22.3	227	10	Q8RZB3
20	59.5	22.3	652	5	Q96418
21	59.5	22.3	901	16	Q9KE04
22	59	22.3	250	5	Q9NM45
23	59	22.3	278	16	Q8YPN8
24	59	22.1	335	16	Q83180
25	58.5	22.1	194	5	P91787
26	58.5	21.9	317	16	Q8DFT3
27	58.5	21.9	585	17	Q8TH08
28	58.5	21.9	589	16	Q9A357
29	58.5	21.9	609	10	Q9CAF1
30	58.5	22.1	657	16	Q9J0J5
31	58.5	21.9	676	5	Q9W1Q6
32	58.5	21.9	840	10	Q9LPP2
33	58	21.7	205	16	Q9KPI1
34	58	21.7	347	17	Q59074
35	58	21.9	573	5	Q19663
36	58	21.9	818	5	Q26059
37	58	21.9	866	16	Q8X9P6
38	58	21.9	5157	3	Q01135
39	57.5	21.7	661	16	Q9JZF3
40	57	21.3	307	16	Q06723
41	57	21.5	611	10	Q8S502
42	57	21.5	619	10	Q8CAA5
43	57	21.3	894	2	Q8VME8
44	56.5	21.2	163	10	Q8S0H1
45	56.5	21.2	177	16	Q9AB84

ALIGNMENTS

RESULT 1

ID	Q93RP0	PRELIMINARY:	PRT: 2523 AA.
AC	Q93RP0;		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)		
DE	XptAl protein.		
GN	XPTAL		
OS	Xenorhabdus nematophilus.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Xenorhabdus.		
OK	NCBI_TaxID=628;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PMFI296;		
RC	MEDLINE=21218513; PubMed=11319082;		
RA	Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;		
RT	"Sequence analysis of insecticidal genes from Xenorhabdus nematophilus		
RT	PMFI296."		
RL	Appl. Environ. Microbiol. 67:2062-2069 (2001).		
DR	EMBL: AJ308438; CAC38401.1; -		
DR	InterPro: IPR000953; Chromo.		
DR	SMART: SM00298; CHROMO: 1.		
SO	SEQUENCE 2523 AA; 286999 MW; 3159852E0655B5B1 CRC64;		

```

Alignment Scores:
Pred. No.: 2,77e-24 Length: 2523
Score: 233.00 Matches: 45
Percent Similarity: 90.20% Conservatives: 1
Best Local Similarity: 88.24% Mismatches: 5
Query Match: 89.92% Indels: 0
DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x Q93RP0 (1-2523)
QY 1 CAACACATTCGTGATCATATATCAGCATATAAATCGGCAACTCGTCAACGCGGCGCTTG 60
DB 1932 GlnThrMetArgHisHisTyrGlnHisLysMetLeuGlnLeuArgGlnArgAlaLeu 1951
QY 61 CGCGGAAAGCTACTGCAAAATCGTTAACGCGTTTGTCTCTCCATCAGATAAACAATAA 120
DB 1952 ProThrLysArgThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnHisLeuLys 1971
QY 121 CTGCAAGTTACTGCGAGACGTTAGCACACGC 153
DB 1972 LeuGlnGlyTyrTrpGlnThrLeuThrGlnArg 1982

RESULT 2
Q9F9Z3
ID Q9F9Z3 PRELIMINARY; PRT; 2376 AA.
AC Q9F9Z3
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE SepA.
GN SEPA.
OS Serratia entomophila.
OG Plasmid pADAP.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=42906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AIM02;
RX MEDLINE=20416224; PubMed=10960097;
RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
RT "Plasmid-located pathogenicity determinants of Serratia entomophila,
RT the causal agent of amber disease of grass grub, show similarity to
RT the insecticidal toxins of Photobacterium luminescens.";
RL J. Bacteriol. 182:5127-5138(2000).
DR EMBL; AF135182; AAC03642.1; -.
KW Plasmid.
SQ SEQUENCE 2376 AA; 262649 MW; E418DACE22DBB7BF CRC64;

Alignment Scores:
Pred. No.: 1.73e-10 Length: 2376
Score: 137.00 Matches: 28
Percent Similarity: 66.67% Conservatives: 6
Best Local Similarity: 54.90% Mismatches: 17
Query Match: 51.70% Indels: 0
DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x Q9F9Z3 (1-2376)
QY 1 CAACACATTCGTGATCATATATCAGCATATAAATCGGCAACTCGTCAACGCGGCGCTTG 60
DB 1785 GluValThrArgArgAspTyrGlnGluAlaLeuAlaValArgLeuValProAla 1804
QY 61 CGCGGAAAGCTACTGCAAAATCGTTAACGCGTTTGTCTCTCCATCAGATAAACAATAA 120
DB 1805 ProGlnThrArgThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnHisLeuLys 1824
QY 121 CTGCAAGTTACTGCGAGACGTTAGCACACGC 153
DB 1825 LeuLysGlyTyrTrpGlnThrLeuThrGlnArg 1835

RESULT 3

```

```

085160
ID 085160 PRELIMINARY; PRT; 2504 AA.
AC 085160
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Insecticidal toxin complex protein TcBA.
GN TCBA.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., firench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";
RL Science 280:2129-2132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA firench-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF047457; AAC38627.1; -.
DR EMBL; AF346498; AAL18460.1; -.
DR InterPro; IPR000566; Lipogln_cytFABP.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN.1.
SQ SEQUENCE 2504 AA; 280654 MW; 8F29B1693D1047CE CRC64;

Alignment Scores:
Pred. No.: 5,64e-09 Length: 2504
Score: 126.50 Matches: 28
Percent Similarity: 64.71% Conservatives: 5
Best Local Similarity: 54.90% Mismatches: 17
Query Match: 47.74% Indels: 1
DB: 2 Gaps: 1

US-09-856-221-2 (1-154) x 085160 (1-2504)
QY 1 CAACACATTCGTGATCATATATCAGCATATAAATCGGCAACTCGTCAACGCGGCGCTTG 60
DB 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929
QY 61 CGCGGAAAGCTACTGCAAAATCGTTAACGCGTTTGTCTCTCCATCAGATAAACAATAA 120
DB 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnHisLeuLys 1949
QY 121 CTGCAAGTTACTGCGAGACGTTAGCACACGC 153
DB 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960

RESULT 4
Q9RN43
ID Q9RN43 PRELIMINARY; PRT; 2516 AA.
AC Q9RN43
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Toxin A (Toxin complex protein).
GN TCDA.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX Merlo D.J., Wegrich L.M., Roberts J.L., Petell J.K.;

```



```

DE ABC transporter, substrate binding protein.
GN ATU4154 OR AGR-L1399.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
NCBI_TaxID=176299;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=21608550; PubMed=11743193;
RX Wood D.W., Setubal J.C., Kaul R., Monks D.E., Almeida N.F. Jr., Woo L.,
Okura V.K., Zhou Y., Chen L., Wood G.E., Eisen J.A., Karp P.D., Bovee D. Sr.,
Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
Chapman P., Choudhury J., Deatherage G., Gillet W., Grant C.,
Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
Raymond C., Rouse G., Saenphumachak C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58."
RN Science 294:2317-2323(2001).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattling S., Miller N., Blanchard M.,
Roulois B., Gordon B., Cao Y., Askenazi M., Hailing C., Mullin L.,
Houmlel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58."
RN Science 294:2323-2328(2001).
DR EMBL; AE009345; AAL44954.1; ALT_INIT.
DR EMBL; AE008268; AAK83280.1; -.
KW Complete proteome.
SQ SEQUENCE 338 AA; 35571 MW; 879AA0BEFDD3C34F CRC64;

Alignment Scores:
Pred. No.: 1.68 Length: 338
Score: 66.50 Matches: 20
Percent Similarity: 54.17% Conservatve: 6
Best Local Similarity: 41.67% Mismatches: 15
Query Match: 24.91% Indels: 7
DB: 16 Gaps: 3

US-09-856-221-2 (1-154) x Q8U8D8 (1-338)
QY 142 AGCTGTCGAGTAACATTTGCAGTTTGTGTTATCTGAGAGAGCAACAGCGGTTAAGC 83
Db 163 ThrSerAlaHisAsnLeuLeuValAlaAlaLeuGluLysAsnGlyLeuLys---LeuSer 181
QY 82 AATTTCAG-----TACGTTTCGCGCAAGCGCGGTTGAGCAGATTGCGCGCATTT 29
Db 182 AspIleAspValThrTyrLeuSerProAla-----AspAlaAlaAlaPhe 197
QY 28 TATGCTGATGATCAGCAATGG 5
Db 198 AlaserAspIleAspAlatrp 205
QY 28 TATGCTGATGATCAGCAATGG 5
Db 198 AlaserAspIleAspAlatrp 205

RESULT 8
Q9HW74
AC Q9HW74 PRELIMINARY; PRT; 268 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein PA4327.
GN PA4327.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.

```

```

OX NCBI_TaxID=287;
RN SEQUENCE FROM N.A.
RP MEDLINE=20437337; PubMed=10984043;
RX Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Huftnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.H., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen."
RN Nature 406:959-964(2000).
RL EMBL; AE004849; AAG07715.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 268 AA; 30572 MW; 0FE62528619AB9A4 CRC64;

Alignment Scores:
Pred. No.: 2.65 Length: 268
Score: 65.00 Matches: 14
Percent Similarity: 59.46% Conservatve: 8
Best Local Similarity: 37.84% Mismatches: 15
Query Match: 24.53% Indels: 0
DB: 16 Gaps: 0

US-09-856-221-2 (1-154) x Q9HW74 (1-268)
QY 13 GATCATTTATCAGTAAATCGCGCAATCGCTCAACGCGCGGCTTGGCGGCAACGCT 72
Db 132 GluHisAlaGlnAlaArgLeuLeuAlaMetArgGlnArgAlaArgGluProValGlnLys 151
QY 73 ACTGCAATTCGTTAACCCTTGTCTTCTTCACAGATAACAAACAAACCTG 123
Db 152 SerAlaArgGluLeuLeuAlaThrTriPlysProGlnGlnAsnLeuAspLeu 168

RESULT 9
Q9VGC3
ID Q9VGC3 PRELIMINARY; PRT; 886 AA.
AC Q9VGC3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG4848 protein (GH23636P).
GN CG4848.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daves L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foshier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

```

```
Alignment Scores:
Pred. No.:      6.96      Length:      1550
Score:          63.00     Matches:      15
```

[illegible]

Dd		717	ASPVALLYSNLSINLEIIRGLYLeuValLeuRbysLeuAspArgaspLeuEunILF	739
Qy	130	TACTGGCAGACGTTAGCACAA	150	
			:::	
Dd	737	SerTrpValAlaLeuAlaGlu	743	
RESULT	11			
Q8YOD5				

AD	08Y0D5	PRELIMINARY	PRI	275 AA
AC	08Y0D5			
DT	01-MAR-2002	(TREMBLrel. 20, Created)		
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)		
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE		Hypothetical protein All3898.		
DE		ALL3898.		
GN		Anabaena sp. (strain PCC 7120).		
OS		Anabaena sp. (strain PCC 7120).		
OC		Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.		
OX		NCBI TaxID:103690;		

RN	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=21595285; PubMed=11759840;
RX	Kaneko T., Nakamura Y., Wolk C.P.,
RA	Kuritz T., Sasamoto S.

RA Watanabe A., Iriduchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuo A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium *Anabaena* sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003594; BAB75597.1; -.

DR InterPro; IPR0000073; Anhydrolase.
DR InterPro; IPR003089; AB_hydrolase.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00561; abhydrolase; 1.

DR PRINTS; PROUILLI; ABHYDROLASE.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 275 AA; 31078 MW; 404A151E935CA7FE CRC64;

Alignment Scores:

Pred. No.:	7.19	Length:	275
Score:	62.00	Matches:	17

Percent Similarity: 65.00% Conservative: 9
 Best Local Similarity: 42.50% Mismatches: 10
 Query Match: 23.40% Indels: 4
 DB: 16 Gaps: 2

US-09-856-221-2 (1-154) x Q8YQD5 (1-275)

QY 28 AAAATGCGGCAACTGGTCAACGGCGCCTTGCAGCGGAA--CGTACTGCAAT 81
 Db 138 LysMetArgGlnLeuileLysArgSerProLeuLeuPheLysileMetArgSerLeuSer 157
 QY 82 TCGTACCCCTTGTCTTCTCCAGTAAACAAAACTGCAAGTACTGCGAGCG 141
 Db 158 PheLeuThrLysilePhe-----GlyLeuAspLysileLysileGlnAspTrpGlnThr 175

RESULT 12
 Q9R2H9 ID Q9R2H9 PRELIMINARY; PRT; 339 AA.
 AC Q9R2H9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE putative TrbE (Fragment).
 GN ORF10.
 OS Pseudomonas putida.
 OG Plasmid pM3.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 ON NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20431897; PubMed=10974112;
 RA Greated A., Tikok M., Krasowiak R., Fairclough R.J., Thomas C.M.;
 RT "The replication and stable-inheritance functions of IncP-9 plasmid
 RT pM3.";
 RL Microbiology 146:2249-2258(2000).
 DR EMBL; AF078924; AAD46131.2; -;
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 339 AA; 38957 MW; 036E9E8A411E848 CRC64;

Alignment Scores:
 Pred. No.: 7.46 Length: 339
 Score: 62.00 Matches: 16
 Percent Similarity: 53.66% Conservative: 6
 Best Local Similarity: 39.02% Mismatches: 19
 Query Match: 23.22% Indels: 0
 DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x Q9R2H9 (1-339)

QY 145 CTACGCTGCGCAGTAACTTTCGATTTTGTGTTATCTGAGGAAGCAACGCGTTA 86
 Db 185 LeuTyrlsAlaSerAsnLeuValLeuArgSerLeuAlaAspTyrglyThrArgArgLeu 204
 QY 85 AGCAATTGACGATACGTTTCGCGCGCAGCGCGGTTGCGGCGATTTTAT 26
 Db 205 ArgValGlnArgHisProGluProArgMetGlyArgValAspArgGluThrAlaPheAsn 224

QY 25 GCT 23
 Db 225 Ala 225

RESULT 13
 Q9R37 ID Q9R37 PRELIMINARY; PRT; 108 AA.
 AC Q9R37;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE Hypothetical 12.0 kDa protein.
 GN MDV082 OR MDV102.
 OS Turkey herpesvirus.

Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Marek's disease-like viruses.
 OX NCBI_TaxID=10390;
 RW [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD5;
 RX MEDLINE=20392152; PubMed=10933706;
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
 RT "The genome of a very virulent Marek's disease virus.";
 RL J. Virol. 74:7980-7988(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD5;
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF243438; AAG14285.1; -;
 DR EMBL; AF243438; AAG14274.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 108 AA; 12031 MW; 67DD21868D980AD7 CRC64;

Alignment Scores:
 Pred. No.: 8.53 Length: 108
 Score: 61.00 Matches: 16
 Percent Similarity: 48.78% Conservative: 4
 Best Local Similarity: 39.02% Mismatches: 20
 Query Match: 22.85% Indels: 1
 DB: 12 Gaps: 0

US-09-856-221-2 (1-154) x Q9DH37 (1-108)

QY 125 TGCAGTTTTTTGTTATCTGAGGAAGCAACGCGTTTACGAATTTGCAGTACGTTTC 66
 Db 16 CysSerPheLeu-ArgHisAspSerGlySerThrGlnAlaValAsnAspThrTyrrValas 35
 QY 65 GCCGCGAAGCGCGCGTTCGACGAGTTCGCCGATTTTATGCTGATGATGATCACCAGT 6
 Db 35 pArgAlaArgProSerAlaAspAlaLysGluHisCysAlaAlaSerAspProGluGluTr 55

QY 5 G 5
 Db 55 p 55

RESULT 14
 Q8QXW5 ID Q8QXW5 PRELIMINARY; PRT; 1557 AA.
 AC Q8QXW5;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 175.7 kDa protein.
 OS Garlic virus E.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Allexivirus.
 ON NCBI_TaxID=150285;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VH;
 RX MEDLINE=21578618; PubMed=11722009;
 RA Chen J., Chen J.P., Adams M.J.;
 RT "Molecular characterisation of a complex mixture of viruses in garlic
 RT with mosaic symptoms in China.";
 RL Arch. Virol. 146:1841-1853(2001).
 DR EMBL; AJ292230; CAC83705.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 1557 AA; 175705 MW; FC62457A811CED74 CRC64;

Alignment Scores:
 Pred. No.: 13.5 Length: 1557
 Score: 61.00 Matches: 20
 Percent Similarity: 42.67% Conservative: 12
 Best Local Similarity: 26.67% Mismatches: 15
 Query Match: 23.02% Indels: 28
 DB: 12 Gaps: 3

US-09-856-221-2 (1-154) x Q8QXW5 (1-1557)		
QY	10	CGTCATCATTCACGATAAAATCGCGCAA- : : :
Db	673	ArgAsnProHISAlaHisPheLeuArgGlnThrAspAlaGlnThrThrHisGlnGln 692
QY	40	-----CTGCCTCAACGCGCG- : : :
Db	693	GlnPheValAspMetAlaHisGlyLeuLysArgGlyCysArgGlyThrLysValAspAla 712
QY	67	AAACCTACTGCA-
Db	713	LysArgAlaThrAlaPheMetSerAspValLysAsnAsnLeuThrGlyLeuValLeuPro 732
QY	106	CAGATAACAAAAATCGCAAGCTTACTGCGCAGCGTTACACAA 150
Db	733	LysLeuAspArgAspLeuSerSerTirpValAlaLeuAlaGlu 747
RESULT 15		
Q9P696	Q9P696	
ID	Q9P696	PRELIMINARY; PRT; 269 AA.
AC	Q9P696;	
DT	01-OCT-2000	(TREMBLrel. 15, Created)
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	Hypothetical 29.4 kDa protein.	
GN	B1D1.250.	
OS	Neurospora crassa.	
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
OC	Sordariales; Sordariaceae; Neurospora.	
OX	NCBI:taxid=5141;	
RP	[1]	SEQUENCE FROM N.A.
RA	Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,	
RA	Nyakatura G., Naves H.W., Mannhaupt G.;	
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RA	German Neurospora genome project;	
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AL355927; CAB91274.2; -	
KL	Hypothetical protein.	
KW	Hypothetical protein.	
SQ	SEQUENCE 269 AA; 29370 MW; 94947D331E95B7A0 CRC64;	
Alignment Scores:		
Pred. No.:	11.8	Length: 269
Score:	60.50	Matches: 16
Percent Similarity:	56.82%	Conservative: 9
Best Local Similarity:	36.36%	Mismatches: 16
Query Match:	22.83%	Indels: 3
DB:	3	Gaps: 2
US-09-856-221-2 (1-154) x Q9P696 (1-269)		
QY	13	GATCATTCACGATAAAATCGCGCAACTCGCTCAACGCGCGCTTGCG- : : : : :
Db	143	AspHis-----HisGlyArgArgGlyLeuLysGluArgAlaLysLeuProTrpLeuHis 160
QY	70	CGTACTCCAAATTCGTTAACCGCTTGTTCCTTCCTCAGATAACAAAAATCGCAAGT 129
Db	161	GlyThrLysSerGlyLeuGluAlaValPheLeuGluGlyLeuGlnHisGlnGlyAsn 180
QY	130	TACTGCGCAGACG 141
Db	181	GluTrpGlnSer 184

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 15, 2003, 15:39:33 ; Search time 26.94 Seconds
(without alignments)
2878.683 Million cell updates/sec

Title: US-09-856-221-3
Perfect score: 507
Sequence: 1 aataccttgctcaacttac.....tacggccctgcatatggtgc 291

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlp
-MODEL=frame+n2p-model -DEV-xlp
-O=/cgn2_1/USPTO.spool/US09856221/runat_15012003_153924_12438/app.query.fasta_1.1948
-DB-A_Geneseq_101002 -QPMF=fastan -SUFFIX=rag -MINWATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAVY=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221@cgn_1_60 -runat_15012003_153924_12438 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGESUB -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRADE=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPEXT=6 -FGAPEXT=7
-YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	293	57.8	2376	22	AA97694	SepA protein encoded by Phototransduc lum
2	220	43.4	579	18	AAW18301	Toxin TcdAIII, encoded by Phototransduc lum
3	220	43.4	579	19	AAW56574	Toxin TcdA, encoded by Phototransduc lum
4	220	43.4	2516	18	AAW17899	Toxin TcdA, encoded by Phototransduc lum
5	220	43.4	2516	19	AAW56572	Toxin TcdA, encoded by Phototransduc lum
6	220	43.4	2516	22	AAW22609	Toxin TcdB, encoded by Phototransduc lum
7	219	43.2	2522	20	AA973729	Streptococcus poly
8	216	42.6	2517	22	AAW22611	Streptococcus poly
9	216	42.6	2537	22	AAW22614	Streptococcus poly
10	208	41.0	573	18	AAW18303	Novel human diagnostic
11	208	41.0	573	19	AAW56559	Synthetic TcdAIII
12	208	41.0	2504	18	AAW17871	Wheat magnesium ch
13	208	41.0	2504	19	AAW56557	Plasmodium falciparum
14	208	41.0	2504	22	AAW22610	Rice magnesium ch
15	208	41.0	2505	22	AAW22612	Human G protein-coupled
16	109	21.5	1565	18	AAW18305	Corynebacterium glutamicum
17	109	21.5	1565	18	AAW56568	Corynebacterium glutamicum
18	77	15.2	294	23	ABP30152	Bacterial general
19	77	15.2	295	23	ABP28317	Drosophila melanogaster
20	71	14.0	296	23	ABP48318	Pneumocystis carinii
21	69.5	13.0	419	22	ABG17301	Novel human diagnostic
22	69	12.9	288	22	ABG25746	Novel human diagnostic
23	68	13.4	20	19	AAW56591	Synthetic TcdAIII
24	67	13.2	1031	22	AAW22610	Phototransduc lum
25	67	12.5	1345	21	AAW18284	Wheat magnesium ch
26	66	13.0	1118	22	AAW48264	Plasmodium falciparum
27	65.5	12.2	186	23	AAU80553	Rice magnesium ch
28	65.5	12.9	700	22	AAW79448	Human G protein-coupled
29	65.5	12.9	706	22	AAW22655	Corynebacterium glutamicum
30	65.5	12.9	706	22	AAW79447	C glutamicum prote
31	65	12.8	299	20	AAW22568	Corynebacterium glutamicum
32	64.5	12.7	1002	22	ABW59634	Bacterial general
33	64	12.6	464	22	ABW61748	Drosophila melanogaster
34	64	12.6	978	22	AAW17509	Pneumocystis carinii
35	63.5	11.9	868	22	ABG22938	Novel human diagnostic
36	63	11.8	140	21	AAW05359	Arabidopsis thaliana
37	63	11.8	165	21	AAW05358	Arabidopsis thaliana
38	63	11.8	191	21	AAW05357	Arabidopsis thaliana
39	63	12.4	243	18	AAW5254	H. pylori ORF 05qp
40	63	11.8	338	22	ABW68636	Drosophila melanogaster
41	63	12.4	350	18	AAW55563	H. pylori ORF 08ce
42	63	11.8	719	13	AAW25069	mLIF-R Mus musculus
43	63	11.8	719	15	AAW49507	Human LIF-R clone
44	63	11.8	719	15	AAW45775	Murine leukaemia 1
45	63	11.8	719	16	AAW474096	Murine leukaemia 1

ALIGNMENTS

RESULT 1
AA97694
ID AA97694 standard; Protein; 2376 AA.
XX
XX AA97694;
AC
XX
XX
DT 19-JUN-2001 (first entry)
DE
DE SepA protein encoded by Serratia insecticidal protein complex gene.
XX
XX Insecticidal protein complex; amber disease; insect; Coleoptera;
XX
XX pesticide; SepA protein.
XX
XX Serratia sp.
XX
XX
XX WO200116305-A2.
XX
XX
XX 08-MAR-2001.
XX

PF 04-SEP-2000; 2000WO-NZ00174.
 XX 02-SEP-1999; 99NZ-0337610.
 XX (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
 XX Glare TR, Hurst MRH, Jackson TA;
 XX WPI: 2001-169009/17.
 DR N-PSDB; AAA91292.
 XX New nucleic acid encoding a polypeptide useful as a pesticide
 PT especially for Coleoptera -
 XX Claim 24; Page 92-99; 109pp; English.
 XX This sequence represents the SepA protein encoded by the Serratia
 CC insecticidal protein complex gene of the invention. The invention relates
 CC to a gene encoding an insecticidal protein complex or a functional
 CC fragment, a neutral mutation, or a homologue of the complex. The
 CC polypeptides and nucleotides of the invention are used to induce amber
 CC diseases or like conditions in insects, especially to those from the
 CC order comprising Coleoptera, useful as a pesticide.
 XX SQ Sequence 2376 AA;
 Alignment Scores:
 Pred. No.: 4.53e-29 Length: 2376
 Score: 293.00 Matches: 65
 Percent Similarity: 80.21% Conservative: 12
 Best Local Similarity: 67.71% Mismatches: 19
 Query Match: 57.79% Indels: 1
 DB: 22 Gaps: 0
 US-09-856-221-3 (1-291) x AA9Y7694 (1-2376)
 QY 1 AATACCTGCTCAACATTACTGAACGCGAGGATGCGACGACTGGCAGATTCTGCGAA 60
 Db 1907 AsnThrLeuLeuGlyThrGluArgGlnAspAlaGluAlaLeuLeuGluGln 1926
 QY 61 ACTCAAGCGAGTGAATGCTTTGCGACAGATGATTAAATGCGAGC- AAGATGATTGCTGAA 119
 Db 1927 ThrGlnGlySerGluLeuLeuArgGlnGlyLeuArgGlnGlnAspValLeuGluGlu 1946
 QY 120 ATTGATGCTGATGATGCGCTTAAGGAAGCGCTCATGTCATGCGATCTCGTTTGAC 179
 Db 1947 IleAspAlaAspIleAlaLeuGluGluSerArgGlyAlaGlnMetArgPheGlu 1966
 QY 180 AGCTTCAGTACGCTGATGACGAGATGTTAACTCCGGTGAACAAACAGCGATGATCTG 239
 Db 1967 ArgTyrLysValLeuTyrGluAlaAspValAsnThrGlyGluLysGlnAlaMetAspLeu 1986
 QY 240 TATCTCTCTTCATGCGTATTGACACCGACGACGATGCGCCCTCATATG 287
 Db 1987 TyrLeuSerSerValLeuSerAlaSerThrAlaAlaLeuPheLeu 2002
 RESULT 2
 AA18301
 ID AA18301 standard; Protein; 579 AA.
 AC AA18301;
 XX 29-JAN-1998 (first entry)
 XX Photorhabdus luminescens insect toxin TcdA111.
 XX Insecticide; insect; pest control; biological control;
 KW Photorhabdus luminescens; TcdA; Southern corn rootworm;
 KW Colorado potato beetle; Western corn rootworm; meal worm;
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
 KW Diptera, Dictyoptera; Acarina; Homoptera.

XX OS Photorhabdus luminescens strain W-14 (ATCC 55397).
 XX PN W09717432-A1.
 XX PD 15-MAY-1997.
 XX PF 06-NOV-1996; 96WO-US18003.
 XX PR 28-AUG-1996; 96US-0705484.
 PR 06-NOV-1995; 95US-0007255.
 PR 28-FEB-1996; 96US-0608423.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
 PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA;
 XX WPI: 1997-281022/25.
 DR N-PSDB; AA768845.
 XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
 PT can be genetically engineered into insect larvae food and plants for
 PT insect control
 XX Claim 34; Page 213-215; 276pp; English.
 XX This polypeptide comprises the 63.6 kDa TcdA111 insecticidal toxin
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
 CC from a genomic DNA clone (AA768845) and includes N-terminal and
 CC tryptic peptide sequences obtained from the isolated protein.
 CC TcdA111 is a proteolytic cleavage product of TcdA (see AA17899).
 CC Claimed toxins of P. luminescens (see AA17871, AA17884-89,
 CC AA17899-900, AA18301-06) can be produced by recombinant DNA methods and
 CC applied to, or genetically engineered into, insect larvae food and
 CC plants for insect control. The toxins are particularly effective
 CC against Southern corn rootworm, Colorado potato beetle, Western corn
 CC rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet
 CC armyworm, black cutworm, cabbage looper, codling moth, corn earworm,
 CC European corn borer, tobacco hornworm and tobacco budworm
 CC (Lepidoptera), and are also active against insects of the orders
 CC Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All
 CC claimed).
 XX SQ Sequence 579 AA;
 Alignment Scores:
 Pred. No.: 1.02e-19 Length: 579
 Score: 220.00 Matches: 50
 Percent Similarity: 74.19% Conservative: 19
 Best Local Similarity: 53.76% Mismatches: 24
 Query Match: 43.39% Indels: 1
 DB: 18 Gaps: 0
 US-09-856-221-3 (1-291) x AA18301 (1-579)
 QY 1 AATACCTGCTCAACATTACTGAACGCGAGGATGCGACGACTGGCAGATTCTGCGAA 60
 Db 101 SerThrLeuGlnAsnIlelleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuGln 120
 QY 61 ACTCAAGCGAGTGAATGCTTTGCGACAGATGATTAAATGCGAGC- AAGATGATTGCTGAA 119
 Db 121 AsnGlnAlaAlaGluLeuIlelleThrAsnLeuSerIleGlnAspLysThrIleGluGlu 140
 QY 120 ATTGATGCTGATGATGCGCTTAAGGAAGCGCTCATGTCGTCACAAATCTCTTTGAC 179
 Db 141 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlyAlaGlnSerArgPheAsp 160
 QY 180 AGCTTCAGTACGCTGATCAGCAAGATGTTAACTCCGGTGAACAAACAGCGATGATCTG 239
 Db 161 SerTyrGlyLysLeuTyrAspGluAsnIlelleGluArgGlnAlaLeuMetThrLeu 180

XX PN WO9717432-AL.
XX PD 15-MAY-1997.
XX PF 06-NOV-1996; 96WO-US18003.
XX PR 28-AUG-1996; 96US-0705484.
XX PR 06-NOV-1995; 95US-0007255.
XX PR 28-FEB-1996; 96US-0608423.
XX PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
XX PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
XX PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
XX PI Strickland JA;
XX XX WPI; 1997-281022/25.
XX DR N-PSDB; AAT68843.
XX XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
XX PT can be genetically engineered into insect larvae food and plants for
XX PT insect control
XX PS Claim 34; Page 185-194; 276pp; English.
XX CC This polypeptide comprises the 282.9 kDa TcdA insecticidal toxin
XX CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
XX CC from a genomic DNA clone (AAV68843) and includes N-terminal and
XX CC tryptic peptide sequences obtained from the isolated protein. The
XX CC 282.9 kDa insect toxin is proteolytically processed into TcdAII and
XX CC TcdAIII components (see AAW17900 and AAW18301). Claimed toxins of P.
XX CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can
XX CC be produced by recombinant DNA methods and applied to, or genetically
XX CC engineered into, insect larvae food and plants for insect control.
XX CC The toxins are particularly effective against Southern corn rootworm,
XX CC Colorado potato beetle, Western corn rootworm, meal worm, boll weevil,
XX CC and turf grub (coleoptera), beet armyworm, black cutworm, cabbage
XX CC looper, codling moth, corn earworm, European corn borer, tobacco
XX CC hornworm and tobacco budworm (Lepidoptera), and are also active
XX CC against insects of the orders Hymenoptera, Diptera, Dictyoptera,
XX CC Acarina and Homoptera. (All claimed).
XX XX N-PSDB; AAV29928.
XX PT Isolated toxins from Photorhabdus luminescens strains - useful for
XX PT control of insect pests
XX PS Claim 34; Pages 224-231; 321pp; English.
XX CC The present sequence represents a protein named TcdA of the bacterium
XX CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
XX CC nematodes of the Heterorhabditis genus. The bacterium has at least 4
XX CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
XX CC produced from these regions that are associated with insecticidal
XX CC activity. The native toxins are secreted proteins. The proteins are
XX CC toxic to insects upon exposure and especially when ingested. The
XX CC nucleic acid sequence can be used to produce transgenic plants,
XX CC baculoviruses or microbial hosts for toxin production. They can be used
XX CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
XX CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
XX CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
XX CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
XX CC codling moth, corn earworm, European corn borer or tobacco hornworm
XX CC or budworm.
XX SQ Sequence 2516 AA;
Alignment Scores:
Pred. No.: 1,53e-19 Length: 2516
Score: 220.00 Matches: 50
Percent Similarity: 74.19% Conservative: 19
Best Local Similarity: 53.76% Mismatches: 24
Query Match: 43.39% Indels: 1
DB: 18 Gaps: 0
US-09-856-221-3 (1-291) x AAW17899 (1-2516)
QY 1 RATACCTGCTACACTACTACAGCGCAGATGCGACAGCACTGCGAGATTGCTGCAA 60
Db 2038 SerThrLeuGlnAsnIleleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuGln 2057
QY 61 ACTCAAGCGCAGTGAATTACCTTTGCGAGAGTATTAAATGACGCGC-AAGATGATTGCTGAA 119
Db 2058 AsnGlnAlaAlaGluLeuIleuThrAsnLeuSerIleGlnAspIleThrIleGluGlu 2077
QY 120 ATTGATGCTGATGAATGCGCTTAAGGAAGCGCTGATGCGACAAATCTCGTTTTCGAC 179
Db 2078 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlnSerArgPheAsp 2097
QY 180 AGCTTCAGTACGCTGATGACGACGATCTTAACTCCGGTGCAAAACAGCGATGATCGT 239
Db 2098 SerTyrgLysLeuTyrgLysAsnIleAsnAlaGlnAlaGlnAlaMetThrLeu 2117
QY 240 TATCTCTCTTCATCGGTATTGACGACACGACGATGACGCC 278

Db 2118 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2130
RESULT 5
AAW56572
ID AAW56572 standard; Protein; 2516 AA.
XX AC AAW56572;
XX DT 07-AUG-1998 (first entry)
XX DE Toxin TcdA, encoded by the tcdA gene from genomic region tcd.
XX KW Photorhabdus luminescens W-14; nematode; symbiotic;
XX KW Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;
XX KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
XX KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
XX KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
XX KW cabbage looper; codling moth; corn earworm; European corn borer;
XX KW Tobacco hornworm; budworm.
XX OS Photorhabdus luminescens.
XX PN WO9808932-AL.
XX PD 05-MAR-1998.
XX PF 05-MAY-1997; 97WO-US07657.
XX PR 06-NOV-1996; 96WO-US18003.
XX PR 28-AUG-1996; 96US-0705484.
XX PR 06-NOV-1996; 96US-0743699.
XX PA (DOWC) DOWELANCO.
XX PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
XX PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
XX PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
XX PI Strickland JA, Sukhapinda K;
XX DR WPI; 1998-179427/16.
XX DR N-PSDB; AAV29928.
XX PT Isolated toxins from Photorhabdus luminescens strains - useful for
XX PT control of insect pests
XX PS Claim 34; Pages 224-231; 321pp; English.
XX CC The present sequence represents a protein named TcdA of the bacterium
XX CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
XX CC nematodes of the Heterorhabditis genus. The bacterium has at least 4
XX CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
XX CC produced from these regions that are associated with insecticidal
XX CC activity. The native toxins are secreted proteins. The proteins are
XX CC toxic to insects upon exposure and especially when ingested. The
XX CC nucleic acid sequence can be used to produce transgenic plants,
XX CC baculoviruses or microbial hosts for toxin production. They can be used
XX CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
XX CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
XX CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
XX CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
XX CC codling moth, corn earworm, European corn borer or tobacco hornworm
XX CC or budworm.
XX SQ Sequence 2516 AA;
Alignment Scores:
Pred. No.: 1,53e-19 Length: 2516
Score: 220.00 Matches: 50
Percent Similarity: 74.19% Conservative: 19
Best Local Similarity: 53.76% Mismatches: 24
Query Match: 43.39% Indels: 1

CC luminescens rather than the nematodes. The toxins have activity against
 CC Lepidopteran insects such as Cabbage Looper (*Trichoplusia ni*),
 CC European Corn Borer (*Ostrinia nubilalis*) and Fall Armyworm
 CC (*Spodoptera frugiperda*) and also against Coleopteran insects
 CC (e.g., Colorado Potato Beetle, *Leptinotarsa decimlineata*). In
 CC addition the toxins are active against strains resistant to known
 CC insecticides. The DNA sequence can be used to generate transgenic plants
 CC of various species that are resistant to economically important insect
 CC pests and also for recombinant production of the toxins for use as
 CC insecticides.

XX SQ Sequence 2522 AA;

Alignment Scores:
 Pred. No.: 2,07e-19 Length: 2522
 Score: 219.00 Matches: 50
 Percent Similarity: 75.27% Conservative: 20
 Best Local Similarity: 53.76% Mismatches: 23
 Query Match: 43.20% Indels: 1
 DB: 20 Gaps: 0

US-09-856-221-3 (1-291) x AAY33729 (1-2522)

QY 1 AATACCTTGCTCAACATCTACTGACAGCGCAGATGCGACAGCTGGCAATGCTGCAA 60
 :||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2044 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuLeuGln 2063
 QY 61 ACTCAAGCGAGTGAATAGCTTTCAGACTATTAAATCGAGGC-AAGATGATGCTGAA 119
 ||| :||||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2064 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 2083
 :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCGCTCATGTGTCACATCTCGTTTGAC 179
 :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2084 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlyAlaGlnSerArgPheAsp 2103
 QY 180 AGCTTCAGTACGCTGATGACAGAGATGTTAACTCCGGTGAAGAAACAGCGATGATCTG 239
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2104 SerTyrSerLysLeuHisAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2123
 QY 240 TATCTCTCTTCATCGGTATGAGCAGCAGCATACGGCC 278
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2124 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2136

RESULT 8
 ID AAB72611 standard; Protein: 2517 AA.
 AC AAB72611;
 XX
 DT 04-MAY-2001 (first entry)
 DE Modified Photorhabdus tcdA toxin.
 KW TcdA; TcdA; insect toxin; plant; insect resistance.
 XX
 OS Photorhabdus sp.
 OS Synthetic.
 XX
 PN WO200111029-A1.
 XX
 PD 15-FEB-2001.
 XX
 PF 11-AUG-2000; 2000WO-US22237.
 XX
 PR 11-AUG-1999; 99US-0148356.
 XX
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 XX
 DR WPI: 2001-191536/19.
 DR N-PSDB; AAF58780.

XX Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm -
 XX
 PS Claim 1; Page 72-83; 106pp; English.
 : XX
 : CC The present invention provides the protein and coding sequences of
 : CC modified versions of the Photorhabdus TcdA and TcdB toxins. These are
 : CC suitable for expression in plants. The toxins are effective against
 : CC insects upon ingestion, and the sequences provided can be used to produce
 : CC transgenic plants with insect resistance. The present sequence is the
 : CC modified Photorhabdus TcdA protein.
 XX SQ Sequence 2517 AA;

Alignment Scores:
 Pred. No.: 5,09e-19 Length: 2517
 Score: 216.00 Matches: 50
 Percent Similarity: 74.19% Conservative: 19
 Best Local Similarity: 53.76% Mismatches: 24
 Query Match: 42.60% Indels: 1
 DB: 22 Gaps: 0

US-09-856-221-3 (1-291) x AAB72611 (1-2517)

QY 1 AATACCTTGCTCAACATCTACTGACAGCGCAGATGCGACAGCTGGCAATGCTGCAA 60
 :||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2039 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuLeuGln 2058
 QY 61 ACTCAAGCGAGTGAATAGCTTTCAGACTATTAAATCGAGGC-AAGATGATGCTGAA 119
 ||| :||||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2059 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 2078
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCGCTCATGTGTCACATCTCGTTTGAC 179
 :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2079 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlyAlaGlnSerArgPheAsp 2098
 QY 180 AGCTTCAGTACGCTGATGACAGAGATGTTAACTCCGGTGAAGAAACAGCGATGATCTG 239
 :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2099 SerTyrGlyLysLeuHisAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2118
 QY 240 TATCTCTCTTCATCGGTATGAGCAGCAGCATACGGCC 278
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 2119 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2131

RESULT 9
 AAB72614
 ID AAB72614 standard; Protein: 2537 AA.
 XX
 AC AAB72614;
 XX
 DT 04-MAY-2001 (first entry)
 DE TcdA toxin-zein ER signal peptide fusion protein.
 XX
 KW TcdA; TcdA; insect toxin; plant; insect resistance.
 XX
 OS Chimeric - Photorhabdus sp.
 OS Chimeric - Zea mays.
 XX
 PN WO200111029-A1.
 XX
 PD 15-FEB-2001.
 XX
 PF 11-AUG-2000; 2000WO-US22237.
 XX
 PR 11-AUG-1999; 99US-0148356.
 XX
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;

CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,
 CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black
 CC cutworm, cabbage looper, codling moth, corn earworm, European corn
 CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and
 CC are also active against insects of the orders Hymenoptera, Diptera
 CC Dictyoptera, Acarina and Homoptera. (All claimed).
 XX

SQ Sequence 2504 AA;

Alignment Scores:
 Pred. No.: 5,61e-18 Length: 2504
 Scores: 208.00 Matches: 46
 Percent Similarity: 72.92% Conservative: 24
 Best Local Similarity: 47.92% Mismatches: 26
 Query Match: 41.03% Indels: 1
 DB: 18 Gaps: 0

US-09-856-221-3 (1-291) x AAW17871 (1-2504)

QY 1 AATACCTTGCTCAACATTACTGAACGCGAGATGACAGACACTGGCAGAAATGCTGCAA 60
 Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuGln 2051
 QY 61 ACTCAAGCGCAGTCAATAGCTTTCAGAGATTAATAATGCAGGC-AGATGATGCTGAA 119
 Db 2052 ThrGlnAlaSerGlnLeuLeuLeuSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071
 QY 120 ATTGATGCTGATGAGTGGCGCTTAAGAAAGCCGTCATGTCGACAAATCTGCTTTGAC 179
 Db 2072 LeuAspSerGlnLeuLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091
 QY 180 AGCTTCAGTACGCTGATGACGAGATGTTAACTCCGGTGAACAAAGCGATGGATCTG 239
 Db 2092 SerTyrSerGlnLeuTyrGluGluAsnAlaGlnAlaGlyGlnArgAlaLeuAlaLeu 2111
 QY 240 TATCTCTCTTCATCGTATTGAGCAGCAGCAGTACGCGCCCTGCATATG 287
 Db 2112 ArgSerGlnSerAlaIleGlnSerGlnGlyAlaGlnIleSerArgMet 2127

RESULT 13

ID AAW56557
 ID AAW56557 standard; Protein: 2504 AA.

XX AAW56557;

AC AAW56557;

DT 07-AUG-1998 (first entry)

DE Toxin TcBa, encoded by the tcba gene from genomic region tcb.

XX Photorhabdus luminescens W-14; nematode; symbiotic;
 KW Heterorhabditis; tcb; tcb; tcb; tcb; insecticidal activity; toxin;
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW Tobacco hornworm; budworm.

XX Photorhabdus luminescens.

OS Photorhabdus luminescens.

PN WO9808932-A1.

XX 05-MAR-1998.

XX 05-MAY-1997; 97WO-US07657.

XX 06-NOV-1996; 96WO-US18003.

XX 28-AUG-1996; 96US-0705484.

XX 06-NOV-1996; 96US-0743699.

XX (DOWC) DOWELANCO.

PA (WISC) WISCONSIN ALDUMI RES FOUND.

XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;

PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 XX Strickland JA, Sukhapinda K;
 DR WPI; 1998-179427/16.
 DR N-PSDB; AAV29985.
 XX Isolated toxins from Photorhabdus luminescens strains - useful for
 PT control of insect pests

XX Claim 34; Pages 163-169; 321pp; English.

XX The present sequence represents a protein named TcBa of the bacterium
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4
 CC distinct genomic regions, tcb, tcb, tcb, and tcb. Peptide products are
 CC produced from these regions that are associated with insecticidal
 CC activity. The native toxins are secreted proteins. The proteins are
 CC toxic to insects upon exposure and especially when ingested. The
 CC nucleic acid sequence can be used to produce transgenic plants,
 CC baculoviruses or microbial hosts for toxin production. They can be used
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 CC codling moth, corn earworm, European corn borer or tobacco hornworm
 CC or budworm.

SQ Sequence 2504 AA;

Alignment Scores:
 Pred. No.: 5,61e-18 Length: 2504
 Scores: 208.00 Matches: 46
 Percent Similarity: 72.92% Conservative: 24
 Best Local Similarity: 47.92% Mismatches: 26
 Query Match: 41.03% Indels: 1
 DB: 19 Gaps: 0

US-09-856-221-3 (1-291) x AAW56557 (1-2504)

QY 1 AATACCTTGCTCAACATTACTGAACGCGAGATGACAGACACTGGCAGAAATGCTGCAA 60

Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuGln 2051

QY 61 ACTCAAGCGCAGTCAATAGCTTTCAGAGATTAATAATGCAGGC-AGATGATGCTGAA 119

Db 2052 ThrGlnAlaSerGlnLeuLeuLeuSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071

QY 120 ATTGATGCTGATGAGTGGCGCTTAAGAAAGCCGTCATGTCGACAAATCTGCTTTGAC 179

Db 2072 LeuAspSerGlnLeuTyrGluGluAsnAlaGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091

QY 180 AGCTTCAGTACGCTGATGACGAGATGTTAACTCCGGTGAACAAAGCGATGGATCTG 239

Db 2092 SerTyrSerGlnLeuTyrGluGluAsnAlaGlnValSerLeuAlaGlyGlnArgAlaLeuAlaLeu 2111

QY 240 TATCTCTCTTCATCGTATTGAGCAGCAGCAGTACGCGCCCTGCATATG 287

Db 2112 ArgSerGlnSerAlaIleGlnSerGlnGlyAlaGlnIleSerArgMet 2127

RESULT 14

ID AAW56557

XX AAW56557 standard; Protein: 2504 AA.

XX AAW56557;

XX 04-MAY-2001 (first entry)

XX Photorhabdus tcbA toxin.

XX TcBa; TcBa; insect toxin; plant; insect resistance.

XX Photorhabdus sp.

XX WO200111029-A1.
 PN Synthetic.
 XX WO200111029-A1.
 PD 15-FEB-2001.
 XX 11-AUG-2000; 2000WO-US22237.
 PF 11-AUG-1999; 99US-0148356.
 XX (DOWC) DOW AGROSCIENCES LLC.
 PA Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 XX WPI: 2001-191536/19.
 DR N-PSDB; AAF58779.
 XX Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm
 XX Disclosure; Page 62-72; 106pp; English.
 XX The present invention provides the protein and coding sequences of
 CC modified versions of the Photobabidus tcba and tcba toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC Photobabidus tcba protein.
 XX SQ Sequence 2504 AA;

Alignment Scores:
 Pred. No.: 5,61e-18 Length: 2504
 Score: 208.00 Matches: 46
 Percent Similarity: 72.92% Conservative: 24
 Best Local Similarity: 47.92% Mismatches: 26
 Query Match: 41.03% Indels: 1
 DB: 22 Gaps: 0

US-09-856-221-3 (1-291) x AAB72610 (1-2504)
 QY 1 AATACCTGCTCAACATTACTGAACGCGAGGATGCGAGCACTGGCAGATTGCTGCAA 60
 Db 2032 SerSerLeuLeuGlyThrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2051
 QY 61 ACTCAAGCGAGTGAATGCTTTCAGAGTATTAAATGCGGC -AAGATGATTGCTGAA 119
 Db 2052 ThrGlnAlaSerGluLeuLeuLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGAAAGCGGTCATGGTGCAACATCTCGTTTGTAC 179
 Db 2072 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091
 QY 180 AGCTTCAGTACGCTGATGACCAAGATGTTAACTCCGGTGAAGAAACAAAGCGATGATCTG 239
 Db 2092 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluGlnArgAlaLeuAlaLeu 2111
 QY 240 TATCTCTCTTCATCGGTATTGACACGACGAGTACGCGCCCTGCATATG 287
 Db 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127

RESULT 15
 AAB72612
 ID AAB72612 standard; Protein; 2505 AA.
 XX AAB72612;
 AC
 XX 04-MAY-2001 (first entry)
 DT
 XX Modified Photobabidus tcba toxin.
 DE
 XX Tcda; Tcda; insect toxin; plant; insect resistance.
 KW

XX Photobabidus sp.
 OS Synthetic.
 XX WO200111029-A1.
 PN 15-FEB-2001.
 XX 11-AUG-2000; 2000WO-US22237.
 PF 11-AUG-1999; 99US-0148356.
 XX (DOWC) DOW AGROSCIENCES LLC.
 PA Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 XX WPI: 2001-191536/19.
 DR N-PSDB; AAF58781.
 XX Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm
 XX Claim 1; Page 83-93; 106pp; English.
 XX The present invention provides the protein and coding sequences of
 CC modified versions of the Photobabidus tcba and tcba toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC modified Photobabidus tcba protein.
 XX SQ Sequence 2505 AA;

Alignment Scores:
 Pred. No.: 5,61e-18 Length: 2505
 Score: 208.00 Matches: 46
 Percent Similarity: 72.92% Conservative: 24
 Best Local Similarity: 47.92% Mismatches: 26
 Query Match: 41.03% Indels: 1
 DB: 22 Gaps: 0

US-09-856-221-3 (1-291) x AAB72612 (1-2505)
 QY 1 AATACCTGCTCAACATTACTGAACGCGAGGATGCGAGCACTGGCAGATTGCTGCAA 60
 Db 2033 SerSerLeuLeuGlyThrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2052
 QY 61 ACTCAAGCGAGTGAATGCTTTCAGAGTATTAAATGCGGC -AAGATGATTGCTGAA 119
 Db 2053 ThrGlnAlaSerGluLeuLeuLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2072
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGAAAGCGGTCATGGTGCAACATCTCGTTTGTAC 179
 Db 2073 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2092
 QY 180 AGCTTCAGTACGCTGATGACCAAGATGTTAACTCCGGTGAAGAAACAAAGCGATGATCTG 239
 Db 2093 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluGlnArgAlaLeuAlaLeu 2112
 QY 240 TATCTCTCTTCATCGGTATTGACACGACGAGTACGCGCCCTGCATATG 287
 Db 2113 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2128

Search completed: January 15, 2003, 15:50:37
 Job time : 35.94 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:45:23 : Search time 9.37561 Seconds
(without alignments)
1826.456 Million cell updates/sec

Title: us-09-856-221-3
Perfect score: 507
Sequence: 1 aatccttgcatactac.....tacggccctgcattatggtgc 291

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+_n2p.model -DEV=xlip
-O=/cgn2_1/USPTO.spool/US09856221/runat_15012003_153926_12521/app_query.fasta_1.1948
-DB-Issued Patents_AA -QWFF=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOFCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221@cgn_1_1_24@runat_15012003_153926_12521 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/6C.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	43.2	2522	4	US-09-251-645-13
2	65	12.8	299	4	US-09-222-938A-34
3	64	12.6	978	2	US-08-415-593-43
4	63	11.8	719	1	US-07-943-843-4
5	63	11.8	719	1	US-08-347-003-4
6	61.5	11.5	799	4	US-09-165-396-4
7	60.5	11.3	152	6	5187153-4
8	59	11.6	100	3	US-09-034-916-13
9	58.5	10.9	169	1	US-08-741-406-2
10	58.5	10.9	169	4	US-09-024-472-2
11	58.5	10.9	782	4	US-09-543-084A-32
12	58.5	10.9	782	4	US-09-543-084A-33

c 13	58.5	10.9	782	4	US-09-543-084A-34	Sequence 34, Appl
c 14	58.5	10.9	782	4	US-09-543-084A-35	Sequence 35, Appl
c 15	58.5	10.9	782	4	US-09-543-084A-36	Sequence 36, Appl
c 16	58	11.4	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
c 17	57.5	10.7	218	5	PCT-US94-14106-57	Sequence 57, Appl
c 18	57.5	10.7	218	5	PCT-US94-14106-61	Sequence 61, Appl
c 19	57	11.2	147	4	US-09-347-833-10	Sequence 10, Appl
c 20	57	10.7	366	3	US-08-875-811-55	Sequence 55, Appl
c 21	57	10.7	750	4	US-09-185-160-14	Sequence 14, Appl
c 22	56.5	11.1	1164	4	US-09-457-708-2	Sequence 2, Appl
c 23	56	10.5	75	1	US-08-289-458-2	Sequence 2, Appl
c 24	56	10.5	75	1	US-08-761-549-2	Sequence 2, Appl
c 25	56	10.5	75	1	US-09-127-646-2	Sequence 2, Appl
c 26	56	11.0	244	4	US-09-134-001C-4784	Sequence 4784, Ap
c 27	56	11.0	385	2	US-08-387-942C-24	Sequence 24, Appl
c 28	56	11.0	419	4	US-09-011-197-4	Sequence 4, Appl
c 29	56	10.5	488	2	US-08-928-692-10	Sequence 10, Appl
c 30	56	10.5	488	4	US-09-339-972-10	Sequence 10, Appl
c 31	56	11.0	496	2	US-08-949-637-2	Sequence 2, Appl
c 32	56	11.0	496	4	US-09-291-488-2	Sequence 2, Appl
c 33	56	11.0	696	1	US-08-765-081-5	Sequence 5, Appl
c 34	56	11.0	696	3	US-09-098-082-5	Sequence 5, Appl
c 35	56	11.0	696	5	PCT-US95-06994-5	Sequence 5, Appl
c 36	56	11.0	703	5	PCT-US95-06994-8	Sequence 8, Appl
c 37	56	11.0	1403	2	US-08-387-942C-3	Sequence 3, Appl
c 38	55.5	10.4	94	2	US-08-828-008-2	Sequence 2, Appl
c 39	55.5	10.4	246	4	US-09-185-160-7	Sequence 7, Appl
c 40	55.5	10.4	2186	2	US-08-822-445-2	Sequence 2, Appl
c 41	55.5	10.4	2186	4	US-09-396-540-2	Sequence 2, Appl
c 42	55	10.8	385	2	US-08-387-942C-26	Sequence 26, Appl
c 43	55	10.8	482	4	US-09-430-564-2	Sequence 2, Appl
c 44	55	10.8	482	4	US-09-430-564-3	Sequence 3, Appl
c 45	55	10.8	997	2	US-08-387-942C-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-251-645-13
; Sequence 13, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Wartha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-13

Alignment Scores:
Pred. No.: 5.38e-23 Length: 2522
Score: 219.00 Matches: 50
Percent Similarity: 75.27% Conservative: 20
Best Local Similarity: 53.76% Mismatches: 23
Query Match: 43.20% Indels: 1
DB: 4 Gaps: 0

US-09-856-221-3 (1-291) x US-09-251-645-13 (1-2522)

QY 1 AATACCTTCTCAACATTACTGAACGCAGGATGCGAAGCACTGGCAGCAATTCGTGCAA 60

```

; Patent No. 5912140
; Patent No. 5912140 5776726
; GENERAL INFORMATION:
; APPLICANT: Whoriskey, Susan K.
; APPLICANT: Quinn, Cheryl L.
; APPLICANT: Tao, Nialan
; APPLICANT: Politis-Virk, Karen I.
; APPLICANT: Schimmel, Paul R.
; TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA
; TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/415,593
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPI94-09
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-862-9540
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 978 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-415-593-43

```

RESULT 2

```

US-09-222-938A-34
; Sequence 34, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Phillip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-222-938A-34

```

```

Alignment Scores:
Pred. No.: 0.768 Length: 299
Score: 65.00 Matches: 25
Percent Similarity: 45.5% Conservative: 16
Best Local Similarity: 27.7% Mismatches: 39
Query Match: 12.8% Indels: 10
DB: 4 Gaps: 2

```

```
US-09-856-221-3 (1-291) x US-09-222-938A-34 (1-299)
```

```

QY 36 AGAAGCACTGGCAGAAATTCCTCAAACTCAAGGCAGTGAATTAGCTTTGACAGATATTA 95
Db 176 GlnSerMetAsnGluIleAsnAlaAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 195
QY 96 AATCGACGCAAGATGATTCCTCAAAATTCATGCTGATGATGATGATGATGATGATGAT 155
Db 196 GluAlaAspLysIleLysIleValThrAlaAlaGluAlaGluAlaGluLysAspArgLeu 215
QY 156 CATGGT-----GCACAAATCTCGT-----TTTGACAGCTTC 185
Db 216 HisGlyValGlyIleAlaGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 235
QY 186 AGTAGCTGTATGACCAAGATGTTAACTCCGGTGAACCAAGCAAGCATGATCTGATCTC 245
Db 236 ThrGluLeuLysGluAlaAsnValGlyMetThrGluGluGlnIleMetSerIleLeuLeu 255
QY 246 TCTTCATCGGTATGACCAAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 275
Db 256 ThrAsnGlnTyrLeuAspThrLeuAsnThr 265

```

RESULT 3

```

US-08-415-593-43
; Sequence 43, Application US/08415593

```

```

; Patent No. 5912140
; Patent No. 5912140 5776726
; GENERAL INFORMATION:
; APPLICANT: Whoriskey, Susan K.
; APPLICANT: Quinn, Cheryl L.
; APPLICANT: Tao, Nialan
; APPLICANT: Politis-Virk, Karen I.
; APPLICANT: Schimmel, Paul R.
; TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA
; TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/415,593
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPI94-09
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-862-9540
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 978 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-415-593-43

```

Alignment Scores:

Pred. No.:	1.75	Length:	978
Score:	64.00	Matches:	29
Percent Similarity:	49.44%	Conservative:	15
Best Local Similarity:	32.58%	Mismatches:	33
Query Match:	12.62%	Indels:	12
DB:	2	Gaps:	3

```

US-09-856-221-3 (1-291) x US-08-415-593-43 (1-978)
QY 13 AACATTACTCAAGCGCAGGATGTCAGAACGACTGCAGAA---TTGCTGCAAACTCAAGCG 69
Db 455 AsnPheMetGluLysPheProAlaAspPheIleAlaGluGlyLeuValGlnThrArgGly 474
QY 70 AGTCAATTACCTTTGCACAGATATTAAATCCAGCGAAGATGATTCCTGAAATTCATGCTG 129
Db 475 TrpPheTyrThrLeuLeu-----ValLeuGlyValGlnLeu 486
QY 130 ATGAAGTGGCGCTTAAAGGAAAGCGCTCATGGTCACAAATCTCGTTTTCACAGCTTCAGTA 189
Db 487 PheGlyIleAlaProPheLysAsnValIleValAsnGlyLeuValLeuAlaSerAspGly 506
QY 190 CGCTGTATGACGAAGATGTTAACTCCGGTGAACCAAGCAAGCATGATCTGATCTCTCTT 249
Db 507 LysLysMetSerLysArgLeu-----LysAsnTyrProGluLeuSerIleValLeu 523
QY 250 CATCGGTATTGACGACCAAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 276
Db 524 AspLysTyrGlyAlaAspAlaLeuArg 532

```

RESULT 4

US-07-943-843-4
 ; Sequence 4, Application US/07943843
 ; Patent No. 5284755
 ; GENERAL INFORMATION:
 ; APPLICANT: Gearing, David P.
 ; TITLE OF INVENTION: Leukemia Inhibitory Factor Receptors
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Immunex Corporation
 ; STREET: 51 University Street
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/943,843
 ; FILING DATE: 19920911
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seese, Kathryn A.
 ; REGISTRATION NUMBER: 32,172
 ; REFERENCE/DOCKET NUMBER: 2606-B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-587-0430
 ; TELEFAX: 206-587-0606
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 719 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-943-843-4

Alignment Scores:
 Pred. No.: 2.16 Length: 719
 Score: 63.00 Matches: 27
 Percent Similarity: 38.78% Conservative: 11
 Best Local Similarity: 27.55% Mismatches: 30
 Query Match: 11.78% Indels: 30
 DB: 1 Gaps: 5

US-09-856-221-3 (1-291) x US-07-943-843-4 (1-719)

QY 290 CACCATATGCGGCGGTACT-----GCTGGTCTCAATACCGATGAGA 246
 Db 378 HisArgilegluglyLeuthrasnGluThrTyArgLeuGlyValGlnMetHisProGly 397
 QY 245 GAGATACAGATCATCGCTGTGTTTTCACCGAGTTAAC----- 207
 Db 398 Gln---GlutleHisasnPhetrLeuthrGlyArgAsnProLeuGlyGlnAlaGlnSer 416
 QY 206 ATCTCTGTCATACAGCTGACTGAGCTGTC----- 177
 Db 417 AlavalVallleasnValThrGluargValAlaProHisAspProThrSerLeuLysVal 436
 QY 176 -----AAACAGAGATTGTCACCATGACGGCTTCTTAAGCGCCACTTCATCAGC 126
 Db 437 LysAspIleasnSerThrValThrPheSerTrpTrpLeuProGlyAsnPhetrLys 456
 QY 125 ATCAATTTCAGCAATCATCTTCCTCGCATTTTAAATACTCTGCAAGCTAATTC 72
 Db 457 IleAsn-----LeuLeuCysGlnleGlutleCysLysAlaAsnSer 470

RESULT 5
 US-08-347-003-4
 ; Sequence 4, Application US/08347003
 ; Patent No. 5785967

; GENERAL INFORMATION:
 ; APPLICANT: Gearing, David P.
 ; APPLICANT: Beckmann, M. P.
 ; TITLE OF INVENTION: Leukemia Inhibitory Factor Receptors
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Immunex Corporation
 ; STREET: 51 University Street
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/347,003
 ; FILING DATE: 29-NOV-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/943,843
 ; FILING DATE: 11-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seese, Kathryn A.
 ; REGISTRATION NUMBER: 32,172
 ; REFERENCE/DOCKET NUMBER: 2606-B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-587-0430
 ; TELEFAX: 206-587-0606
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 719 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-347-003-4

Alignment Scores:
 Pred. No.: 2.16 Length: 719
 Score: 63.00 Matches: 27
 Percent Similarity: 38.78% Conservative: 11
 Best Local Similarity: 27.55% Mismatches: 30
 Query Match: 11.78% Indels: 30
 DB: 1 Gaps: 5

US-09-856-221-3 (1-291) x US-08-347-003-4 (1-719)

QY 290 CACCATATGCGGCGGTACT-----GCTGGTCTCAATACCGATGAGA 246
 Db 378 HisArgilegluglyLeuthrasnGluThrTyArgLeuGlyValGlnMetHisProGly 397
 QY 245 GAGATACAGATCATCGCTGTGTTTTCACCGAGTTAAC----- 207
 Db 398 Gln---GlutleHisasnPhetrLeuthrGlyArgAsnProLeuGlyGlnAlaGlnSer 416
 QY 206 ATCTCTGTCATACAGCTGACTGAGCTGTC----- 177
 Db 417 AlavalVallleasnValThrGluargValAlaProHisAspProThrSerLeuLysVal 436
 QY 176 -----AAACAGAGATTGTCACCATGACGGCTTCTTAAGCGCCACTTCATCAGC 126
 Db 437 LysAspIleasnSerThrValThrPheSerTrpTrpLeuProGlyAsnPhetrLys 456
 QY 125 ATCAATTTCAGCAATCATCTTCCTCGCATTTTAAATACTCTGCAAGCTAATTC 72
 Db 457 IleAsn-----LeuLeuCysGlnleGlutleCysLysAlaAsnSer 470

RESULT 6
 US-09-165-396-4
 ; Sequence 4, Application US/09165396
 ; Patent No. 6441134

```

; GENERAL INFORMATION:
; APPLICANT: BECKER, JEFFREY W.
; APPLICANT: LUBKOWITZ, MARK A.
; TITLE OF INVENTION: ISOLATED CANDIDA ALBICANS OLIGOPEPTIDE TRANSPORTER GENE
; FILE REFERENCE: 372.6520P
; CURRENT APPLICATION NUMBER: US/09/165,396
; EARLIER FILING DATE: 1998-10-02
; EARLIER APPLICATION NUMBER: PCT/US98/02332
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: 60/037,859
; EARLIER FILING DATE: 1997-02-07
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 799
; TYPE: PRT
; ORGANISM: S. cerevisiae
; US-09-165-396-4

Alignment Scores:
Pred. No.: 3,74 Length: 799
Score: 61.50 Matches: 21
Percent Similarity: 44.44% Conservative: 15
Best Local Similarity: 25.93% Mismatches: 36
Query Match: 11.50% Indels: 9
DB: 4 Gaps: 3

US-09-856-221-3 (1-291) x US-09-165-396-4 (1-799)
QY 271 CTGCTGGTGCATCAATACCGATGAGAGATACAGATCCATCGCTGTTTTCACCG--- 215
Db 373 LeuYrPheThrAsnThrTrpTyAlaLysTyMetProValIleSerGlySerThrTyr 392
QY 214 GAGTTACACTTCGTCATACAGCGTACTGAAGCTGTCGAACGAGATGTGCACCATGA 155
Db 393 AspAsnThrGlnAsnLysTyAsnValThrLysIleLeuAsnGluAsp----- 408
QY 154 CGGCTTCTTAAAGCGCCACTTCATCAGCATCAATTTACGAACTATTCGCTGCAATTT 95
Db 409 --TyrSerIleAsnLeuGluLysTyLysLysTySerProValPheValPro----- 425
QY 94 TAATACCTGCAAGAACTAATTCGCTGCTGAGTTTCGACCAATTCGCAATGCTGCTG 35
Db 426 ---PheSerTyLeuLeuSerTyAlaLeuAsnPheAlaValIleAlaValPheVal 444
QY 34 CAT 32
Db 445 His 445

RESULT 7
5187153-4
; Patent No. 5187153
; APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
; TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S
; AMYLOID POLYPEPTIDE DERIVATIVES
; NUMBER OF SEQUENCES: 33
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/502,273
; FILING DATE: 29-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 361,912
; FILING DATE: 06-JUN-1989
; APPLICATION NUMBER: 359,911
; FILING DATE: 12-MAY-1989
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO: 4

; GENERAL INFORMATION:
; APPLICANT: BECKER, JEFFREY W.
; APPLICANT: LUBKOWITZ, MARK A.
; TITLE OF INVENTION: ISOLATED CANDIDA ALBICANS OLIGOPEPTIDE TRANSPORTER GENE
; FILE REFERENCE: 372.6520P
; CURRENT APPLICATION NUMBER: US/09/165,396
; EARLIER FILING DATE: 1998-10-02
; EARLIER APPLICATION NUMBER: PCT/US98/02332
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: 60/037,859
; EARLIER FILING DATE: 1997-02-07
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 799
; TYPE: PRT
; ORGANISM: S. cerevisiae
; US-09-165-396-4

Alignment Scores:
Pred. No.: 3,74 Length: 799
Score: 61.50 Matches: 21
Percent Similarity: 44.44% Conservative: 15
Best Local Similarity: 25.93% Mismatches: 36
Query Match: 11.50% Indels: 9
DB: 4 Gaps: 3

US-09-856-221-3 (1-291) x 5187153-4 (1-152)
QY 233 CATCGCTTGTTC---ACCGAGTTAAATCTTCCTCATACACGCTACGAGCTGTC 177
Db 12 TyrArgValPhePheLysSerHisLeuSerIlePhePheLeuGlnAlaLysGlu---Val 30
QY 176 AAACAGAGATTGTGC---ACCATGACGCGCTTCCTTAAGCGCCACTTCATCAGCATCAAT 120
Db 31 LeuGlyArgLeuCysLeuThrValLeuLeuPheMetSerLys----- 44
QY 119 TTCAGCAATCATCTTCCTGCGATTTCATCTCTGCAAAAGCTAATTC 72
Db 45 ---ThrAsnTrpLeuSerCysIleLeu---LeuCysAsnThrGlySer 58

; RESULT 8
; US-09-034-916-13
; Sequence 13, Application US/09034916
; Patent No. 6046314
; GENERAL INFORMATION:
; APPLICANT: GEBE, JOHN A.
; APPLICANT: SIADAK, ANTHONY W.
; APPLICANT: ARUFFO, ALEJANDRO A.
; TITLE OF INVENTION: SCALPHA: A NOVEL SCAVENGER RECEPTOR
; TITLE OF INVENTION: CYSTEINE-RICH DOMAIN-CONTAINING POLYPEPTIDE AND MONOCLONAL
; TITLE OF INVENTION: ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBINS & ASSOCIATES
; STREET: 90 MIDDLEFIELD ROAD, SUITE 200
; CITY: MENLO PARK
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,916
; FILING DATE: 04-MAR-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,956
; FILING DATE: 06-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION/DOCKET NUMBER: 33,208
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 325-7812
; TELEFAX: (650) 325-7823
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-034-916-13
; Alignment Scores:

```



```

Percent Similarity: 41.12% Conservative: 15
Best Local Similarity: 27.10% Mismatches: 38
Query Match: 10.93% Indels: 25
DB: 4 Gaps: 5

US-09-856-221-3 (1-291) x US-09-024-472-2 (1-169)
QY 286 ATATGACGGCGCTACTGCTGCTCAATACCGATGACGAGATACAGA----- 236
Db 24 IleArgAsnAlaValProLeuGlyThrAlaLysGluGluMetGluArgPheTrpAsn 43
QY 235 -----CCATCCCTGTTTTCACCGAGTTACATCTTCGTATACAGGTA 188
Db 44 LysAsnIleGlySerAsnArgProLeuSerProHisIleThrIleTyrSerTrpSerLeu 63
QY 187 -----CTGAAGCTGTCAAAAGAGATTGTCACCATGACGGCTTTCCTTAAGCGCC 137
Db 64 ProMetAlaMetSerIleCysHisArgGlyThrGly-----IleAlaLeuSerAla 80
QY 136 ---ACTTCATCAGCATCAATTTTCAGCAATCATCTTCGCT----- 101
Db 81 GlyValSerLeuPheGlyMetSerAlaLeuLeuLeuProGlyAsnPheLeuSerTyrLeu 100
QY 100 -----CCATTTAACTCTGCAAAAGCTAATTCATCTGCTTACGTTTCAGCA 53
Db 101 GluLeuValLysSerLeuGlyLeuGlyProAlaLeuIleHisThrAlaLysPheAlaLeu 120
QY 52 ATTCTGCGAGTGTCTTCGAT 32
Db 121 ValPheProLeuMetTyrHis 127

RESULT 11
US-09-543-084A-32
: Sequence 32, Application US/09543084A
: Patent No. 6361988
: GENERAL INFORMATION:
: APPLICANT: Frances H. Arnold
: APPLICANT: Zhixin Shao
: APPLICANT: Huimin Zhao
: APPLICANT: Lorraine J. Giver
: TITLE OF INVENTION: Recombination of Polynucleotide
: TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
: STREET: 2029 Century Park East, Suite 3800
: CITY: Los Angeles
: STATE: CA
: COUNTRY: USA
: ZIP: 90067
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Windows 98
: SOFTWARE: Word 97
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/543,084A
: FILING DATE: April 4, 2000
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/381,935
: APPLICATION NUMBER: 60/041,666
: FILING DATE: March 25, 1997
: APPLICATION NUMBER: 60/045,211
: FILING DATE: April 30, 1997
: APPLICATION NUMBER: 60/046,256
: FILING DATE: May 12, 1997
: APPLICATION NUMBER: 08/905,359
: FILING DATE: August 4, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Oldenkamp, David J.
: REGISTRATION NUMBER: 29,421
: REFERENCE/DOCKET NUMBER: 330187-89

```

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (310) 788-5000
: TELEFAX: (310) 788-5100
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 782 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-09-543-084A-32
Alignment Scores: 10.2 Length: 782
Pred. No.: 58.50 Matches: 17
Score: 50.00% Conservative: 5
Best Local Similarity: 38.64% Mismatches: 17
Query Match: 10.93% Indels: 5
DB: 4 Gaps: 1

US-09-856-221-3 (1-291) x US-09-543-084A-32 (1-782)
QY 229 GCTGTGTTTTCACCGAGTTAACATCTTCGTATACAGCGTACTGAAGCTGTCAAAACGA 170
Db 440 AlaProPheGlnProLeuTyrAlaSerSerGlyGlnAlaValLeuAspGlySerArgSer 459
QY 169 GATTGTCCACCATGACGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTTACGCAATC 110
Db 460 AspCysAla-----LeuGlyAlaAspProAspAlaAlaValProGlyIle 474
QY 109 ATCTTGCTGCTCA 98
Db 475 LeuGlyProAla 478

RESULT 12
US-09-543-084A-33
: Sequence 33, Application US/09543084A
: Patent No. 6361988
: GENERAL INFORMATION:
: APPLICANT: Frances H. Arnold
: APPLICANT: Zhixin Shao
: APPLICANT: Huimin Zhao
: APPLICANT: Lorraine J. Giver
: TITLE OF INVENTION: Recombination of Polynucleotide
: TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
: STREET: 2029 Century Park East, Suite 3800
: CITY: Los Angeles
: STATE: CA
: COUNTRY: USA
: ZIP: 90067
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Windows 98
: SOFTWARE: Word 97
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/543,084A
: FILING DATE: April 4, 2000
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/381,935
: APPLICATION NUMBER: 60/041,666
: FILING DATE: March 25, 1997
: APPLICATION NUMBER: 60/045,211
: FILING DATE: April 30, 1997
: APPLICATION NUMBER: 60/046,256
: FILING DATE: May 12, 1997
: APPLICATION NUMBER: 08/905,359
: FILING DATE: August 4, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Oldenkamp, David J.

```


ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
TELEPHONE: (310) 788-5000
TELEFAX: (310) 788-5100
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-543-084A-33

Alignment Scores:
Pred. No.: 10.2 Length: 782
Score: 58.50 Matches: 17
Percent Similarity: 50.00% Conservative: 5
Best Local Similarity: 38.64% Mismatches: 17
Query Match: 10.93% Indels: 5
DB: 4 Gaps: 1

US-09-856-221-3 (1-291) x US-09-543-084A-33 (1-782)

QY 229 GCTGTTTTCACCGGAGTTAAACATCTTCATACAGCGTGAAGCTGTCAAAACGA 170
DB 440 AlaProPheGlnProLeuTyraIaSerSerGlyGlnAlaValLeuAspGlySerArgSer 459
QY 169 GATTGTCACCATGACGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTTTCAGCAATC 110
DB 460 AspCysAla-----LeuGlyAlaAspProAspAlaAlaValProGlyIle 474
QY 109 ATCTTGCTCTGCA 98
DB 475 LeuGlyProAla 478

RESULT 13

US-09-543-084A-34
Sequence 34, Application US/09543084A
Patent No. 6361986

GENERAL INFORMATION:
APPLICANT: Frances H. Arnold
APPLICANT: Zhixin Shao
APPLICANT: Huimin Zhao
APPLICANT: Lorraine J. Giver
TITLE OF INVENTION: Recombination of Polynucleotide
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 98
SOFTWARE: Word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/543,084A
FILING DATE: April 4, 2000
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/381,935
APPLICATION NUMBER: 60/041,666
FILING DATE: March 25, 1997
APPLICATION NUMBER: 60/045,211
FILING DATE: April 30, 1997
APPLICATION NUMBER: 60/046,256
FILING DATE: May 12, 1997
APPLICATION NUMBER: 08/905,359
FILING DATE: August 4, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
TELEPHONE: (310) 788-5000
TELEFAX: (310) 788-5100
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-543-084A-34

Alignment Scores:
Pred. No.: 10.2 Length: 782
Score: 58.50 Matches: 17
Percent Similarity: 50.00% Conservative: 5
Best Local Similarity: 38.64% Mismatches: 17
Query Match: 10.93% Indels: 5
DB: 4 Gaps: 1

US-09-856-221-3 (1-291) x US-09-543-084A-34 (1-782)

QY 229 GCTGTTTTCACCGGAGTTAAACATCTTCATACAGCGTGAAGCTGTCAAAACGA 170
DB 440 AlaProPheGlnProLeuTyraIaSerSerGlyGlnAlaValLeuAspGlySerArgSer 459
QY 169 GATTGTCACCATGACGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTTTCAGCAATC 110
DB 460 AspCysAla-----LeuGlyAlaAspProAspAlaAlaValProGlyIle 474
QY 109 ATCTTGCTCTGCA 98
DB 475 LeuGlyProAla 478

RESULT 14

US-09-543-084A-35
Sequence 35, Application US/09543084A
Patent No. 6361988

GENERAL INFORMATION:
APPLICANT: Frances H. Arnold
APPLICANT: Zhixin Shao
APPLICANT: Huimin Zhao
APPLICANT: Lorraine J. Giver
TITLE OF INVENTION: Recombination of Polynucleotide
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 98
SOFTWARE: Word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/543,084A
FILING DATE: April 4, 2000
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/381,935
APPLICATION NUMBER: 60/041,666
FILING DATE: March 25, 1997
APPLICATION NUMBER: 60/045,211
FILING DATE: April 30, 1997
APPLICATION NUMBER: 60/046,256
FILING DATE: May 12, 1997

APPLICATION NUMBER: 08/905,359
FILING DATE: August 4, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
REFERENCE/DOCKET NUMBER: 330187-89
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 788-5100
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-543-084A-35

Alignment Scores:
Pred. No.: 10.2 Length: 782
Score: 58.50 Matches: 17
Percent Similarity: 50.00% Conservative: 5
Best Local Similarity: 38.64% Mismatches: 17
Query Match: 10.93% Indels: 5
DB: 4 Gaps: 1

US-09-856-221-3 (1-291) x US-09-543-084A-35 (1-782)
QY 229 GCTGTGTTTTCACCGAGTTAACTTCGTGTCATACAGCGTACTGAAGCTGTCAAAACGA 170
Db 440 AlaProPheGlnProLeuTyAlaSerSerGlyGlnAlaValLeuAspGlySerArgSer 459
QY 169 GATTGTGCACATGACGGCTTTCCCTTAAGCCGCACTTCATCAGCATCAATTTCAGCAATC 110
Db 460 AspCysAla-----LeuGlyAlaAspProAspAlaAlaValProGlyIle 474
QY 109 ATCTTGCTGCA 98
Db 475 LeuGlyProAla 478

RESULT 15
US-09-543-084A-36
Sequence 36, Application US/09543084A
Patent No. 6361988
GENERAL INFORMATION:
APPLICANT: Frances H. Arnold
APPLICANT: Zhixin Shao
APPLICANT: Huimin Zhao
APPLICANT: Lorraine J. Giver
TITLE OF INVENTION: Recombination of Polynucleotide
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 98
SOFTWARE: word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/543,084A
FILING DATE: April 4, 2000
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/381,935
APPLICATION NUMBER: 60/041,666
FILING DATE: March 25, 1997
APPLICATION NUMBER: 60/045,211
FILING DATE: April 30, 1997

APPLICATION NUMBER: 60/046,256
FILING DATE: May 12, 1997
APPLICATION NUMBER: 08/905,359
FILING DATE: August 4, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
REFERENCE/DOCKET NUMBER: 330187-89
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 788-5100
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-543-084A-36

Alignment Scores:
Pred. No.: 10.2 Length: 782
Score: 58.50 Matches: 17
Percent Similarity: 50.00% Conservative: 5
Best Local Similarity: 38.64% Mismatches: 17
Query Match: 10.93% Indels: 5
DB: 4 Gaps: 1

US-09-856-221-3 (1-291) x US-09-543-084A-36 (1-782)
QY 229 GCTGTGTTTTCACCGAGTTAACTTCGTGTCATACAGCGTACTGAAGCTGTCAAAACGA 170
Db 440 AlaProPheGlnProLeuTyAlaSerSerGlyGlnAlaValLeuAspGlySerArgSer 459
QY 169 GATTGTGCACATGACGGCTTTCCCTTAAGCCGCACTTCATCAGCATCAATTTCAGCAATC 110
Db 460 AspCysAla-----LeuGlyAlaAspProAspAlaAlaValProGlyIle 474
QY 109 ATCTTGCTGCA 98
Db 475 LeuGlyProAla 478

Search completed: January 15, 2003, 15:57:47
Job time : 12.3756 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:46:43 : Search time 5.45922 seconds
(without alignments)
2119.222 Million cell updates/sec

Title: US-09-856-221-3
Perfect score: 507
Sequence: 1 ataactgtctcaattac.....tacggccctgcatgtgctc 291

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODE=framet_n2p_model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09856221/runat_15012003_153926_12555/app_query.fasta_1.1948
-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNIT8=bits -START=1 -END=1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=45 -LOCAL=200 -THR_SCORE=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USBR=US09856221 -CGN_1_1_14 -runat_15012003_153926_12555
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	220	43.4	2516	10	US-09-817-514A-2
2	208	41.0	2504	10	US-09-817-514A-8
3	65.5	12.9	706	9	US-09-738-626-6019
4	65	12.8	508	10	US-09-833-745-40

5	62.5	12.3	513	10	US-09-833-745-49	Sequence 49, Appli
6	62	12.2	128	10	US-09-925-299-1486	Sequence 1486, Ap
c 7	61	11.4	296	10	US-09-771-730-116	Sequence 116, App
c 8	61	11.4	296	10	US-09-771-730-120	Sequence 120, App
c 9	61	11.4	303	10	US-09-771-730-118	Sequence 118, App
c 10	61	11.4	332	10	US-09-771-730-28	Sequence 28, Appli
c 11	61	11.4	994	10	US-09-852-909-2	Sequence 2, Appli
12	61	12.0	1319	9	US-10-042-431-14	Sequence 14, Appli
13	61	12.0	1413	9	US-10-042-431-13	Sequence 13, Appli
14	61	12.0	1453	9	US-10-042-431-11	Sequence 11, Appli
15	61	12.0	2167	10	US-09-801-368-56	Sequence 56, Appli
c 16	59.5	11.1	162	10	US-09-747-155-227	Sequence 227, App
17	59	11.6	144	10	US-09-925-301-1538	Sequence 1538, Ap
c 18	59	11.0	313	10	US-09-886-055-461	Sequence 461, App
19	59	11.6	1116	9	US-09-977-577-10	Sequence 10, Appli
20	59	11.6	1149	9	US-09-977-577-11	Sequence 11, Appli
21	59	11.6	1151	9	US-09-977-577-13	Sequence 13, Appli
22	59	11.6	1156	9	US-09-977-577-12	Sequence 12, Appli
c 23	58.5	10.9	787	10	US-09-205-448-8	Sequence 8, Appli
24	58	11.4	350	10	US-09-881-752A-234	Sequence 234, App
25	58	11.4	464	10	US-09-863-824-2	Sequence 2, Appli
c 26	58	10.8	550	9	US-09-738-626-5417	Sequence 5417, Ap
c 27	57.5	10.7	756	10	US-09-919-497-66	Sequence 66, Appli
c 28	57	10.7	56	10	US-09-764-869-1049	Sequence 1049, Ap
29	57	11.2	250	10	US-09-815-242-10871	Sequence 10871, A
c 30	56.5	10.6	94	10	US-09-811-284-203	Sequence 203, App
c 31	56.5	10.6	209	10	US-09-393-634-13	Sequence 13, Appli
c 32	56.5	11.1	241	10	US-09-764-864-936	Sequence 936, App
c 33	56.5	10.6	324	10	US-09-886-055-217	Sequence 217, App
c 34	56.5	10.6	619	9	US-09-895-913A-326	Sequence 326, App
35	56.5	11.1	1164	10	US-09-950-046A-2	Sequence 2, Appli
36	55.5	10.9	94	10	US-09-764-864-1558	Sequence 1558, Ap
c 37	55.5	10.4	94	10	US-09-739-707-2	Sequence 2, Appli
c 38	55.5	10.4	928	9	US-10-108-605-261	Sequence 261, App
c 39	55.5	10.9	969	10	US-09-118-276-2	Sequence 2, Appli
40	55.5	10.9	1354	10	US-09-808-571A-2	Sequence 4, Appli
c 41	55.5	10.9	1447	10	US-09-808-571A-2	Sequence 2, Appli
c 42	55.5	10.4	2186	10	US-09-927-568-2	Sequence 2, Appli
43	55	10.8	378	10	US-09-864-761-43251	Sequence 43251, A
44	55	10.8	469	10	US-09-764-864-1136	Sequence 1136, Ap
c 45	55	10.3	876	9	US-09-738-626-4789	Sequence 4789, Ap

ALIGNMENTS

RESULT 1
US-09-817-514A-2
; Sequence 2, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: firench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent version 3.0
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 2516
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2

Alignment Scores:
Pred. No.: 2.14e-21 Length: 2516
Score: 220.00 Matches: 50
Percent Similarity: 74.19% Conservative: 19
Best Local Similarity: 53.76% Mismatches: 24

```

Query Match: 43.39% Indels: 1
DB: 10 Gaps: 0
US-09-856-221-3 (1-291) x US-09-817-514A-2 (1-2516)

QY 1 AATACCTGCTCAACATTACTCAAGCGCAGGATCCAGACCACTGGCAGAAATTCCTGCAA 60
Db 2038 SerThrLeuGlnAsnIlellelGluArgGlnAspAlaGluAlaLeuAsnAlaLeuLeuGln 2057
QY 61 ACTCAAGCAGTGAATAGCTTTGTCAGAGTATTAAATGCAGGC-AGATGATGCTGAA 119
Db 2058 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspIleThrIleGluGlu 2077
QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCGTCATGTCGCAATCTCGTTTGGAC 179
Db 2078 LeuAspAlaGluysThrValLeuGluysSerIleAlaGlnSerArgPheAsp 2097
QY 180 ACCTTCAGTACGCTGATGACCAAGATGTTAACTCCGGTGAAACAAACAGCGATGATCTG 239
Db 2098 SerTyrglyLysLeuTyrrAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2117
QY 240 TATCTCTCTTCATCGTATTGAGCACCAGCAGTACGCGCC 278
Db 2118 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2130

RESULT 2
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: french-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8

Alignment Scores:
Pred. No.: 9,168-20 Length: 2504
Score: 208.00 Matches: 46
Percent Similarity: 72.92% Conservative: 24
Best Local Similarity: 47.92% Mismatches: 26
Query Match: 41.03% Indels: 1
DB: 10 Gaps: 0

US-09-856-221-3 (1-291) x US-09-817-514A-8 (1-2504)

QY 1 AATACCTGCTCAACATTACTCAAGCGCAGGATCCAGACCACTGGCAGAAATTCCTGCAA 60
Db 2032 SerSerLeuLeuGlytyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2051
QY 61 ACTCAAGCAGTGAATAGCTTTGTCAGAGTATTAAATGCAGGC-AGATGATGCTGAA 119
Db 2052 ThrGlnAlaSerGluLeuIleLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071
QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCGTCATGTCGCAATCTCGTTTGGAC 179
Db 2072 LeuAspSerGluysThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091
QY 180 ACCTTCAGTACGCTGATGACCAAGATGTTAACTCCGGTGAAACAAACAGCGATGATCTG 239
Db 2092 SerTyrglyLysLeuTyrrAspGluAsnIleAsnAlaGlyGluAsnGlnAlaLeu 2111

```

```

QY 240 TATCTCTCTTCATCGTATTGAGCACCAGCAGTACGCGCCCTGCATATG 287
Db 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127

RESULT 3
US-09-738-626-6019
; Sequence 6019, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKURO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6019
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6019

Alignment Scores:
Pred. No.: 1,46 Length: 706
Score: 65.30 Matches: 15
Percent Similarity: 48.28% Conservative: 23
Best Local Similarity: 25.86% Mismatches: 23
Query Match: 12.92% Indels: 7
DB: 9 Gaps: 1

US-09-856-221-3 (1-291) x US-09-738-626-6019 (1-706)

QY 132 CAAGTGGCGCTTAAGCAAGCGTCATGTCGCACATCTCGTTTGACAGCTTCAGTACG 191
Db 335 GluThrAlaGlnSerGluSerValHisGlyThrGluProAspArgAspGluLeuThrMet 354
QY 192 CTGTAT-----GACGAAGATGTTAACTCCGGTGAACAAACAGCG 230
Db 355 PheTyrrMetThrLeuGlnTrpLeuCysAspGluGlnLeuAlaAlaGlnLysArgAla 374
QY 231 ATGATCTGTCATCTCTTCATCGTATTGAGCACCAGCAGTACGCGCCCTGCAT 284
Db 375 ValAspAlaGlyMetSerIleGlyIleMetAlaAspLeuAlaValGlyValHis 392

RESULT 4
US-09-833-745-40
; Sequence 40, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13

```

; PRIOR APPLICATION NUMBER: 60/197,770
 ; PRIOR FILING DATE: 2000-04-14
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 40
 ; LENGTH: 508
 ; TYPE: PRT
 ; ORGANISM: Bacillus subtilis
 US-09-833-745-40

Alignment Scores:
 Pred. No.: 1.54 Length: 508
 Score: 65.00 Matches: 22
 Percent Similarity: 51.35% Conservative: 16
 Best Local Similarity: 29.73% Mismatches: 18
 Query Match: 12.82% Indels: 18
 DB: 10 Gaps: 4

US-09-856-221-3 (1-291) x US-09-833-745-40 (1-508)

QY 108 ATGATGCTGAATTCATGCTGATGAGTCGCGCTTAAGGAAAGCGT----- 155
 Db 204 ValValaLarYrIeGluAlaGluLysLeuAlaTyrGlnThrGluArgIleAlaSerLeu 223
 QY 156 -----CATGGTGCACAAATTCCTTTTGACAGCTTCACCTGATGACGAGATGTT 209
 Db 224 ThrIeGluGlyLeuGlnGlyTyrGlnGluInIleAspValAlaGluArgIleAtrgPheTyr 239
 QY 210 AAC-----TCCGGTGAACAAACAGCGATGATCTG-----TATCTC 245
 Db 240 HisLeuAlaArgGlyTyrGlnGluInIleAspValAlaGluArgIleAtrgPheTyrLeu 259
 QY 246 TCTTCATCGGTATTGAGCACCAGCAGTACGCGCTCATATG 287
 Db 260 SerAspSerglyLeuThrThrSerGlnGlyGluLeuArgVal 273

RESULT 5

US-09-833-745-49
 ; Sequence 49, Application US/09833745
 ; Patent No. US20020052038A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROBERTS, JOSEPH
 ; APPLICANT: SETHURAMAN, NATARAJAN
 ; APPLICANT: MACALLISTER, THOMAS
 ; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
 ; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
 ; FILE REFERENCE: 078728/0106
 ; CURRENT APPLICATION NUMBER: US/09/833,745
 ; CURRENT FILING DATE: 2001-04-13
 ; PRIOR APPLICATION NUMBER: 60/197,770
 ; PRIOR FILING DATE: 2000-04-14
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 513
 ; TYPE: PRT
 ; ORGANISM: Bacillus subtilis
 US-09-833-745-49

Alignment Scores:
 Pred. No.: 3.38 Length: 513
 Score: 62.50 Matches: 21
 Percent Similarity: 52.17% Conservative: 15
 Best Local Similarity: 30.43% Mismatches: 14
 Query Match: 12.33% Indels: 19
 DB: 10 Gaps: 4

US-09-856-221-3 (1-291) x US-09-833-745-49 (1-513)

QY 108 ATGATGCTGAATTCATGCTGATGAGTCGCGCTTAAGGAAAGCGT----- 155
 Db 210 ValValaLarYrIeGluAlaGluLysLeuAlaTyrGlnThrGluArgIleAlaSerLeu 229

QY 156 -----CATGGTGCACAAATTCCTTTTGACAGCTTCACCTGATGACGAGATGTT 209
 Db 230 ThrIeGluGlyLeuGlnGlyTyrGlnGluInIleAspValAlaPhe-----AspGluAspIle 245
 QY 210 AAC-----TCCGGTGAACAAACAGCGATGATCTG-----TAT 242
 Db 246 HisLeuAlaLeuArgGlyTyrGlnGluInIleAspValAlaGluArgIleAtrgPheTyr 265
 QY 243 CTCTCTTCATCGGTATTGAGCACCAGC 269
 Db 266 LeuSerAspSerglyLeuThrThrSer 274

RESULT 6

US-09-925-299-1486
 ; Sequence 1486, Application US/09925299
 ; Patent No. US20020055627A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA102
 ; CURRENT APPLICATION NUMBER: US/09/925,299
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05883
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1556
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1486
 ; LENGTH: 126
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (112)
 ; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
 US-09-925-299-1486

Alignment Scores:
 Pred. No.: 2.53 Length: 126
 Score: 62.00 Matches: 10
 Percent Similarity: 73.33% Conservative: 1
 Best Local Similarity: 66.67% Mismatches: 4
 Query Match: 12.23% Indels: 0
 DB: 10 Gaps: 0

US-09-856-221-3 (1-291) x US-09-925-299-1486 (1-126)

QY 5 CCTTGCTCAACATTACTGACGCGCAGGATGCACAGCAGCTGGCAG 49
 Db 34 ProcysSerLeuLeuArgGlyArgValThrSerHisTrpGln 48

RESULT 7

US-09-771-730-116
 ; Sequence 116, Application US/09771730
 ; Patent No. US20020146807A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Prayaga, Sudhirdas K.
 ; APPLICANT: Li, Li
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Spytek, Kimberly Ann
 ; APPLICANT: Tchernev, Vellizar T.
 ; APPLICANT: Vernet, Corine A. M.
 ; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
 ; FILE REFERENCE: 15966-645
 ; CURRENT APPLICATION NUMBER: US/09/771,730
 ; CURRENT FILING DATE: 2001-08-21
 ; PRIOR APPLICATION NUMBER: 60/178,413
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: 60/178,371
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: 60/178,408


```

; APPLICANT: Padigaru, Muralidhara
; APPLICANT: MacDougall, John R.
; APPLICANT: Spytek, Kimberly Ann
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A. M.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-645
; CURRENT APPLICATION NUMBER: US/09/771,730
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/178,413
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,371
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,408
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,409
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/180,634
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/220,516
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/260,290
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 118
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-730-118

```

```

Alignment Scores:
Pred. No.: 4.57 Length: 303
Score: 61.00 Matches: 22
Percent Similarity: 42.86% Conservative: 14
Best Local Similarity: 26.19% Mismatches: 26
Query Match: 11.40% Indels: 23
DB: 3 Gaps: 3

```

```

US-09-856-221-3 (1-291) x US-09-771-730-118 (1-303)
QY 288 CCATATGCGAGCGCG-----TACTGCTGGTCTCAATACCG---ATGAAGAGA 244
Db 167 ProphecysglyProAsnArgValAspTyrIlePheCysAspIleProAlaMetLeuArg 186
QY 243 GA-----TACAGA 236
Db 187 LeuAlaCysAlaAspThrAlaIleAsnGluLeuValThrPheAlaAspIleGlyPheLeu 206
QY 235 TCCATCGCTGTTTTCACCGAGGTTAACATCTTCGTCATACAGCGTACTGAAGCTGTCA 176
Db 207 AlaLeuThrCysPheMetProIleLeuThrSerTyrGlyTyrIleValAlaAlaIleLeu 226
QY 175 AAACGAGATTGTGCACATGACGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTTC 116
Db 227 ArgIleProSerAlaAspGlyArgArgAsnAlaPheSerThrCysAlaAlaHisLeuThr 246
QY 115 GCAATCATCTTG 104
Db 247 ValValIleVal 250

```

```

RESULT 10
US-09-771-730-28
; Sequence 28, Application US/09771730
; Patent No. US20020146807A1
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: MacDougall, John R.
; APPLICANT: Spytek, Kimberly Ann
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A. M.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-645
; CURRENT APPLICATION NUMBER: US/09/771,730
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/178,413
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,371
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,408
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,370
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,406
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,414
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,409
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/180,634
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/220,516
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,408
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,943
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/257,599
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/260,290
; PRIOR FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-730-28

```

```

Alignment Scores:
Pred. No.: 4.7 Length: 332
Score: 61.00 Matches: 22
Percent Similarity: 42.86% Conservative: 14
Best Local Similarity: 26.19% Mismatches: 26
Query Match: 11.40% Indels: 23
DB: 3 Gaps: 3

```

```

US-09-856-221-3 (1-291) x US-09-771-730-28 (1-332)
QY 288 CCATATGCGAGCGCG-----TACTGCTGGTCTCAATACCG---ATGAAGAGA 244
Db 188 ProphecysglyProAsnArgValAspTyrIlePheCysAspIleProAlaMetLeuArg 207
QY 243 GA-----TACAGA 236
Db 208 LeuAlaCysAlaAspThrAlaIleAsnGluLeuValThrPheAlaAspIleGlyPheLeu 227
QY 235 TCCATCGCTGTTTTCACCGAGGTTAACATCTTCGTCATACAGCGTACTGAAGCTGTCA 176
Db 228 AlaLeuThrCysPheMetProIleLeuThrSerTyrGlyTyrIleValAlaAlaIleLeu 247
QY 175 AAACGAGATTGTGCACATGACGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTTC 116

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:44:53 ; Search time 13.4107 Seconds
(without alignments)
4172.064 Million cell updates/sec

Title: US-09-856-221-3
Perfect score: 507
Sequence: 1 aataccttgctcaacattac.....tacggcctgcatagtgtgc 291

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODE=frame+n2p.model -DEV=xlp
-O=/csn2.1/USPTO.spool/US09856221/runat_15012003_153925_12500/app_query.fasta_1.1948
-DB=PIR_73 -OPMT=fastap -SUFFIX=PIR -MINMATCH=0.1 -LOOPT=0 -LOOPTXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221_@CN_1_1_95_@runat_15012003_153925_12500 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	SUMMARIES				Description
	Score	Query %	Match	Length DB ID	
1	69	13.6	1379	2	S37310 protoporphyrin IX
2	68	13.4	299	2	T32094 hypothetical prote
3	68	13.4	329	2	A69791 hypothetical prote
4	68	13.4	1383	2	T07126 magnesium chelatas
5	68	13.4	1693	2	AC3240 helicase, SNF2 fam
6	67	12.5	1345	2	B71608 DNA-directed RNA p
7	67	13.2	1380	2	S474721 protoporphyrin IX
8	66.5	12.4	520	2	S45753 probable membrane
9	66.5	12.4	585	2	T09782 hydroxymethylgluta
10	66	13.0	218	2	A81764 late competence pr
11	65.5	12.2	930	2	T00403 hypothetical prote
12	65	12.1	252	2	T37633 succinate dehydrog
13	65	12.8	274	2	A95452 SPFH domain/Band 7
14	65	12.8	299	2	G98116 conserved hypothet

15	65	12.8	508	1	UFBSSH
16	65	12.8	1382	2	T01789
17	64.5	12.7	843	2	S78372 DNA-directed RNA p
18	64.5	12.7	1002	2	S54252 deep orange protei
19	64.5	12.7	1033	2	A12359 hypothetical prote
20	64.5	12.1	1036	2	T30839 sarco/endoplasmic
21	64.5	12.1	1127	2	T20870 hypothetical prote
22	64	12.6	1561	2	T00248 zinc finger protei
23	63.5	12.5	220	2	E70749 probable hycP prot
24	63.5	12.5	240	1	B64457 uridylylate kinase h
25	63	12.4	186	2	T07977 protoporphyrin IX
26	63	12.4	256	2	D87298 enoyl-CoA hydratase
27	63	12.4	350	2	D71817 cytochrome-c perox
28	63	11.8	1092	2	JX0312 differentiation-st
29	62.5	12.3	2048	2	C84609 hypothetical prote
30	62	11.6	259	2	T46881 succinate dehydrog
31	62	12.2	297	2	H72670 hypothetical prote
32	62	11.6	489	2	H64224 hypothetical prote
33	62	11.6	582	2	T32903 hypothetical prote
34	62	12.2	770	2	H84463 conserved hypothet
35	61.5	12.1	296	2	E83292 ABC transporter, p
36	61.5	11.5	563	2	H95212 hypothetical prote
37	61.5	11.5	563	2	A98077 probable membrane
38	61.5	11.5	799	2	S50773 hypothetical prote
39	61.5	12.1	1422	2	T24212 probable membrane
40	61	12.0	434	2	S53951 beta-galactosidase
41	61	12.0	724	2	T06687 breifidin A-sensit
42	61	11.4	1036	2	T05687 hypothetical prote
43	61	12.0	1076	2	S50536 bud emergence prot
44	61	12.0	2167	2	S50658 NOL1/NOP2/sun faml
45	60.5	11.3	279	2	C81412

ALIGNMENTS

RESULT 1

S37310
protoporphyrin IX magnesium chelate (EC 4.99.1.-) - garden snapdragon
C:Species: Antirrhinum majus (garden snapdragon)
C:Date: 31-Dec-1993 sequence_revision 31-Dec-1993 #text_change 08-Oct-1999
C:Accession: S37310; S35411
R:Hudson, A.; Carpenter, R.; Doyle, S.; Coen, E.S.
EMBO J. 12, 3711-3719, 1993
A>Title: Olive: a key gene required for chlorophyll biosynthesis in Antirrhinum majus
A:Reference number: S37310; MID:94008977; PMID:8404842
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1379 <HUD>
C:Cross-references: EMBL:X73144; NID:g312128; PIDN:CAA51664.1; PID:g312129
C:Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase
C:Keywords: chlorophyll biosynthesis; lyase

Alignment Scores:				
Pred. No.:	4.43	Length:	1379	
Score:	69.00	Matches:	24	
Percent Similarity:	40.96%	Conservative:	10	
Best Local Similarity:	28.92%	Mismatches:	25	
Query Match:	13.61%	Indels:	24	
DB:	2	Gaps:	3	

US-09-856-221-3 (1-291) x S37310 (1-1379)

QY	99	GCAGGCAAGATGATGCTGAATTTGAT-----	125
Db	1102	AlaValLysMetValAlaGluLeuAspGluProValGluGlnAsnPheValArgLysHis	1121
QY	126	-----GCTGATGAAGTGGCGCTTAAGGAAGCGCTGATGTCGACAAATCTCGT	173
Db	1122	AlaLeuGluGlnAlaLysGluLeuGlyVal---GluValArgGluAlaAlaSerArgile	1140
QY	174	TTTGACAGCTTCAGTACCGCTGATGACGAGATGTTAC-----	212

QY 147 -----GAAAGCCGTCATGTCGCACAAATCTCGTTT 176
 DB 1126 AlsergluclnalacinalaLeuGlyValGluValArgGluAlaAlaThrArgLlePhe 1145
 QY 177 GACAGTTCAGTACGCTGTATGACGAGATGTTAAAC----- 212
 DB 1146 SerAsnAlaSerGlySerTyrSerSerAsnIleAsnLeuAlaValGluAsnSerSerTrp 1165
 QY 213 TCCGGTGAAACAAACGATGCTGCTCTCTTCATCGTATTGACACCAAGCAT 272
 DB 1166 AsnAspGluGlnLeuGlnAspMetTyrLeuSerArgLysSerPheAlaPheAspSer 1185
 QY 273 ACGGCC 278
 DB 1186 AspAla 1187
 RESULT 5
 AC3240
 heliCase, SNF2 family [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AC3240
 R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AC3240
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1693 <KUR>
 A:Cross-references: GB:AE008690; PIDN:AAL4637.1; PID:g17744125; GSPDB:GN00189
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu6101
 A:Genome: plasmid
 Alignment Scores:
 Pred. No.: 6.01 Length: 1693
 Score: 68.00 Matches: 19
 Percent Similarity: 56.60% Conservative: 11
 Best Local Similarity: 35.85% Mismatches: 21
 Query Match: 13.41% Indels: 2
 DB: 2 Gaps: 2
 US-09-856-221-3 (1-291) x AC3240 (1-1693)
 QY 111 ATTGCTGAAATTCATGCTGATGAAGTCGCGCTTAAGGAAAGCGTCATGTCGACAACTCT 170
 DB 1038 ilealagluileglyvalaspnilerlevalaspGluAlaGlnGluPheArgLysLeu 1057
 QY 171 CGTWT---GACAGCTTCAGTACGCTGTATGACGAGATGTTAACTCCGCGTAAACAA 227
 DB 1058 SerPheAlaThrAsnMetSerThrLeuLysGlyValAspProAsn---GlySerGlnArg 1076
 QY 228 CCGATGGATGCTGATCTCTCTTCATCGCTATTGAGCACC 266
 DB 1077 AlaTrpAspLeuTyrValLysSerArgPheIleGluThr 1089
 RESULT 6
 B71608
 DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - malaria parasite (Pla
 N:Alternate names: protein PFB0715W
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 20-Sep-1999
 C:Accession: B71608
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: B71608
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1345 <GAR>
 A:Cross-references: GB:AE001413; GB:AE001362; NID:g3845255; PIDN:AAC71932.1; PID:g384
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0715W
 A:Superfamily: DNA-directed RNA polymerase 132K polypeptide
 C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription
 Alignment Scores:
 Pred. No.: 7.92 Length: 1345
 Score: 67.00 Matches: 22
 Percent Similarity: 51.52% Conservative: 12
 Best Local Similarity: 33.33% Mismatches: 21
 Query Match: 12.52% Indels: 11
 DB: 2 Gaps: 2
 US-09-856-221-3 (1-291) x B71608 (1-1345)
 QY 278 GCCGTCCTGCTGCTCAATACCATGAAGAGATACAGATCCATCGCTGTTTTC 219
 DB 1018 GlyGlnThrAlaGlySer-SerAsnSerAsnThrLysTyrGlyThrThrIleValSerSe 1037
 QY 218 A-----CCGGAGTTAACTCTTCGTCATACAGCGTACTGAAGCT 180
 DB 1037 rThrLysAspThrGluIleProThrLeuThrIleSerSerThrAsnValLeuLysG 1057
 QY 179 GTCAAAACGAGATGTCACCATGACGCTTCCCTTAAGCGCCACTTCATCAGCATCAAT 120
 DB 1057 nTyrLysLysAspCysSer-----LeuSerLeuArgSerAsnGluAsnGlyVal 1074
 QY 119 TTCAGCAATCATCTTG 104
 DB 1074 eAspThrValMetLeu 1079
 RESULT 7
 S64721
 protoporphyrin IX magnesium chelatase (EC 4.99.1.1-) Xantha-f precursor - barley
 C:Species: Hordeum vulgare (barley)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
 C:Accession: S64721
 R:Jensen, P.E.; Willows, R.D.; Petersen, B.L.; Voithkecht, U.C.; Stummann, B.M.; Kann
 Mol. Gen. Genet. 250, 383-394, 1996
 A:Title: Structural genes for Mg-chelatase subunits in barley: Xantha-f, -g and -h.
 A:Reference number: S64721; MUID:96186869; PMID:8602155
 A:Accession: S64721
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1380 <JEN>
 A:Cross-references: EMBL:U26916; NID:g861198; PIDN:AAA99721.1; PID:g861199
 C:Genetics:
 A:Gene: Xantha-f
 A:Introns: 488/1; 495/3; 1096/3
 C:Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase
 C:Keywords: chlorophyll biosynthesis; lyase
 F:1-50/Domain: signal sequence #status predicted <SIG>
 F:51-1380/Product: protoporphyrin IX magnesium chelatase chain Xantha-f #status predi

Alignment Scores:
 Pred. No.: 7.93 Length: 1380
 Score: 67.00 Matches: 24
 Percent Similarity: 40.96% Conservative: 10
 Best Local Similarity: 28.52% Mismatches: 25
 Query Match: 13.21% Indels: 24
 DB: 2 Gaps: 3
 US-09-856-221-3 (1-291) x S64721 (1-1380)
 QY 99 GCAGCGAAGATGCTGCTGAAATTTGAT----- 125
 ||| |||||:|||||:|||||:|||||

A:Residues: 1-218 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA97882.1; PID:g16415192; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: comFC

Alignment Scores:
Pred. No.: 9.44 Length: 218
Score: 66.00 Matches: 17
Percent Similarity: 57.63% Conservative: 17
Best Local Similarity: 28.81% Mismatches: 21
Query Match: 13.02% Indels: 4
DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x AB1764 (1-218)

QY 117 GAATTCATGTCGATGAAGTGGCGCTTAAGAAAGCCGTATGTCGACAAATCTCGTTT 176
Db 145 GluLeuAlaLysLysHisSerGluLysGlnSerLysThrLysArgGluArgLeu 164
QY 177 GACAGCTTCAGTACGCTGAT--GACGAAGATGTTAACTCCGTCGTAAGAAACAGCGATG 233
Db 165 GluThrGluGlnValPheTyrValAspGluGluLeuGlySerSerGluLysGluLeuLeu 184
QY 234 -----GACTCTATCTCTTCATCGTATTGACACACGACGACGACGCTG 281
Db 185 LeuPheAspIleTyrThrGlySerThrLeuAsnLeuAlaGlnAlaLeu 203

RESULT 11
T00403
hypothetical protein At2g44900 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T13E15.9
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00403; B84884
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Masc
submitted to the EMBL Data Library, July 1997
A:Description: Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence.
A:Reference number: 214146
A:Accession: T00403
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-930 <ROU>
A:Cross-references: EMBL:AC002388; NID:g3420042; PID:g2344894
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
M.; KOO, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84884
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <STO>
A:Cross-references: GB:AE002093; NID:g2344894; PIDN:AAC31834.1; GSPDB:GN00139
C:Genetics:
A:Gene: AT2g44900; T13E15.9
A:Map position: 2
A:Introns: 460/3; 513/3; 540/3; 584/3; 656/3; 688/2; 709/3; 768/2; 844/2; 872/3

Alignment Scores:
Pred. No.: 12 Length: 930
Score: 65.50 Matches: 30
Percent Similarity: 41.82% Conservative: 16
Best Local Similarity: 27.27% Mismatches: 36
Query Match: 12.24% Indels: 28
DB: 2 Gaps: 3

US-09-856-221-3 (1-291) x T00403 (1-930)

QY 262 CTCATACCGATGAAGAG-----AGATACAGATCCATCGCT 227
||||| |||||||

Db 223 LeuAsnIleAspGluGluAlaLeuGlyLysValSerValArgTyrLeuSerValAla 242
QY 226 TGTTTTTCA-----CCGGAGTTA 209
Db 243 GlyThrSerAsnIleLysTyrSerIleAlaSerAsnAsnTyrAspLysLeuProLysLeu 262
QY 208 ACATCTTCGTATCAGCGTACTGAAGCTGTCAAACGAGATTTGTCACCATCAGCGTT 149
Db 263 ThrGly-----LeuAspValSerArgThrAspIleGlyProThrAlaVal 277
QY 148 TCCTTAAGGCGCCACTCATCATCAGCATCAATTTGACCAATCATCTTGCCTGCATTTAAATAC 89
Db 278 SerArgPheLeuThrSerSerGlnSerLeuLysValLeu-CysAlaLeuAsnCysHisVa 297
QY 88 TCTGCAAAAGCTAATTCCTACTGCTTGTGAGTTCGTCAGCAATCTGCCAGCTGCTTCGCACTCT 29
Db 297 LLeuGluGluAspGluSerLeuIleSerTyrAsnArgPheLysGlyLysValLeuLeuAl 317
QY 28 GCGGTTTCAGTATGTTGAGCAAGGTATT 1
Db 317 aLeuPheThrAsnValPheAspGlyLeu 326

RESULT 12
T37633
succinate dehydrogenase (ubiquinone) (EC 1.3.5.1) iron-sulfur protein [similarity] -
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-Jan-2001
C:Accession: T37633; E32394
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: 221733
A:Accession: T37633
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-252

A:Cross-references: EMBL:Z99091; NID:g2398800; PIDN:CA811771.1; PID:g2398807; GSPDB:G
A:Experimental source: strain 972h-; cosmid cl3F5
R:Gould, S.J.; Subramani, S.; Scheffler, I.E.
Proc. Natl. Acad. Sci. U.S.A. 86, 1934-1938, 1989
A:Title: Use of the DNA polymerase chain reaction for homology probing: isolation of
A:Reference number: A32394; MUID:89184541; PMID:2494655
A:Accession: E32394
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: mRNA
A:Residues: 89-228 <GOU>
C:Genetics:
A:Gene: sdh2; SPDB:SPAC13F5.08
A:Map position: 1
C:Superfamily: fumarate reductase iron-sulfur protein; ferredoxin 2[4Fe-4S] homology;
C:Keywords: 2Fe-2S; 3Fe-4S; iron-sulfur protein; metalloprotein; mitochondrio
F:54-96/Domain: ferredoxin [2Fe-2S] homology <FBR1>
F:158-240/Domain: ferredoxin 2[4Fe-4S] homology <FBR2>
F:75,80,83,95/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted
F:165,168,171,232/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
F:175,222,228/Binding site: 3Fe-4S cluster (Cys) (covalent) #status predicted

Alignment Scores:
Pred. No.: 12.7 Length: 252
Score: 65.00 Matches: 22
Percent Similarity: 46.77% Conservative: 7
Best Local Similarity: 35.48% Mismatches: 16
Query Match: 12.15% Indels: 17
DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x T37633 (1-252)

QY 258 ATACCATGACAGATACAGATCCATCGCTGTTTTCACCGAGTTAACATCTCTCT 199
Db 59 IleLysIleLysAsnGluInAspProThrLeuThrPheArgArg----- 73
QY 198 CATACACGCTACTGAAGCTGTCAAACGAGATGTGCACCATGACGCTTCTCCTTAAGCG 139
Db 74 -----SerCysArgGluGlyIle-CysGlySerCysAla----- 84

QY 138 CCACCTTCATGACCAATTTTCAGCAATCATCTTGCTGCTGCAATTTAAATACCTCTGCAAGC 79
A:Accession: G98116
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <KUR>
A:Cross-references: GB:AE007317; PIDN:AA00764.1; PID:gl5459662; GSPDB:GN00174
C:Genetics:
A:Gene: spr1962

Db 85 -----MetAsnIleAsnGlySerAsnThrLeuAlaCysIleCysAsnIleLysLysAs 102
QY 78 TAAT 75
Db 102 pAsn 103

RESULT 13
A95252
SPFH domain/Band 7 family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
A:Accession: A95252
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95252
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76210.1; PID:gl4973667; GSPDB:GN00164; TIGR:SP4
A:Experimental source: Strain TIGR4
C:Genetics:
A:Gene: SP2156

Alignment Scores:
Pred. No.: 12.9 Length: 274
Score: 65.00 Matches: 25
Percent Similarity: 45.56% Conservatives: 16
Best Local Similarity: 27.78% Mismatches: 39
Query Match: 12.82% Indels: 10
DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x A95252 (1-274)

QY 36 AGAAGCACTGGCAGAAATCTGCTGCAAACTCAAGCGAGTGAATAGCTTTGCAGAGTATTAA 95
Db 151 GlnSerMetAsnGluIleAsnAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 170
QY 96 AATGCGAGCAAGATGATCTGCTGAAATTCGCTGATGAGTGGCGCTTAAAGGAAGCCGT 155
Db 171 GluAlaAspLysIleLysIleValThrAlaAlaGluAlaGluLysAspArgLeu 190
QY 156 CATGGT-----GCACATCTCGT-----TTTGACAGCTTC 185
Db 191 HisGlyValGlyIleAlaGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 210
QY 186 AGTACGCTGATGACGAAGATGTTAACTCCGGTGAAGAAACAGCGATGATCTGTATCTC 245
Db 211 ThrGluLeuLysGluAlaAsnValGlyMetThrGluGluGluIleMetSerIleLeuLeu 230
QY 246 TCTTCATCGGTATTGAGCAGCAGCAGTACG 275
Db 231 ThrAsnGlnThrLeuAspThrLeuAsnThr 240

RESULT 14
G98116
conserved hypothetical protein spr1962 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
A:Accession: G98116
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; B
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.W.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G98116
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <KUR>
A:Cross-references: GB:AE007317; PIDN:AA00764.1; PID:gl5459662; GSPDB:GN00174
C:Genetics:
A:Gene: spr1962

Alignment Scores:
Pred. No.: 12.9 Length: 299
Score: 65.00 Matches: 25
Percent Similarity: 45.56% Conservatives: 16
Best Local Similarity: 27.78% Mismatches: 39
Query Match: 12.82% Indels: 10
DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x G98116 (1-299)

QY 36 AGAAGCACTGGCAGAAATCTGCTGCAAACTCAAGCGAGTGAATAGCTTTGCAGAGTATTAA 95
Db 176 GlnSerMetAsnGluIleAsnAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 195
QY 96 AATGCGAGCAAGATGATCTGCTGAAATTCGCTGATGAGTGGCGCTTAAAGGAAGCCGT 155
Db 196 GluAlaAspLysIleLysIleValThrAlaAlaGluAlaGluLysAspArgLeu 215
QY 156 CATGGT-----GCACATCTCGT-----TTTGACAGCTTC 185
Db 216 HisGlyValGlyIleAlaGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 235
QY 186 AGTACGCTGATGACGAAGATGTTAACTCCGGTGAAGAAACAGCGATGATCTGTATCTC 245
Db 236 ThrGluLeuLysGluAlaAsnValGlyMetThrGluGluGluIleMetSerIleLeuLeu 255
QY 246 TCTTCATCGGTATTGAGCAGCAGCAGTACG 275
Db 256 ThrAsnGlnThrLeuAspThrLeuAsnThr 265

RESULT 15
UFBSHS
histidine ammonia-lyase (BC 4.3.1.3) huth [similarity] - Bacillus subtilis
N:Alternate names: histidase huth
C:Species: Bacillus subtilis
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
A:Accession: S18810; C69643; T47089
R:Oda, M.; Sugishita, A.; Furukawa, K.
J. Bacteriol. 170, 3199-3205, 1988
A:Title: Cloning and nucleotide sequences of histidase and regulatory genes in the Ba
A:Reference number: S18808; MUID:88257040; PMID:2454913
A:Accession: S18810
A:Molecule type: DNA
A:Residues: 1-508 <ODA>
A:Cross-references: EMBL:M20659; NID:gl43074; PIDN:AAA22538.1; PID:gl43076
R:Experimental source: strain 1A270
A:Authors: F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: C69643
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-508 <XUN>
A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB15971.1; PID:g2636481
A:Experimental source: strain 168
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 143, 337-343, 1995
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contain
A:Reference number: 224350; MUID:95219088; PMID:7704263
A:Accession: T47089
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-508 <YOS>
A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BA06644.1; PID:g603770
A:Experimental source: strain BGSC1A1
C:Genetics:
A:Gene: hutH
C:Function:
A:Description: catalyzes the formation of (E)-3-(1H-imidazol-4-yl)-propenoic acid (uroce
C:Superfamily: Bacillus subtilis histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; histidine catabolism
F:141-143/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:142/Modified site: dehydroalanine (Ser) #status predicted

Alignment Scores:	Pred. No.:	Length:
Score:	13.3	508
Percent Similarity:	65.00	Matches: 22
Best Local Similarity:	51.35%	Conservative: 16
Query Match:	29.73%	Mismatches: 18
DB:	12.82%	Indels: 18
	1	Gaps: 4

US-09-856-221-3 (1-291) x UFBHS (1-508)

QY	108	ATGATTGCTGAATGATGCTGATGAAGTGGCGCTTAAGGAAGCCGT-----	155
Db	204	ValValAlaTyrIleGluAlaGluIleAlaTyrGlnThrGluArgIleAlaSerLeu	223
QY	156	-----CATGGTGCACATCTCGTTTTCACAGCTTCAGTACGCTGTATGACGAAGATGTT	209
Db	224	ThrIleGluGlyLeuGlnGlyIleIleAspAlaPhe-----AspGluAspIle	239
QY	210	AAC-----TCCGGTGAACAAACAGCGATGATCTG-----TATCTC	245
Db	240	HisLeuAlaArgGlyTyrGlnGluGlnIleAspValAlaGluArgIleArgPheTyrLeu	259
QY	246	TCTTCATCGGTATTGACGACCGACGATACGCCCTGCATATG	287
Db	260	SerAspSerGlyLeuThrThrSerGlnGlyGluLeuArgVal	273

Search completed: January 15, 2003, 15:56:24
Job time : 18.4107 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 6.88336 Seconds
(without alignments)
3506.895 Million cell updates/sec

Title: us-09-856-221-3
Perfect score: 507
Sequence: 1 aatacctgtcacaattac.....tacggccctgcatagtgtgc 291

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09856221/runat_15012003_153924_12445/app_query.fasta_1.1948
-DB=SwissProt_40 -QPM=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221.cgn_1.1.38.@runat_15012003_153924_12445 -NCPU=6 -ICPU=3
-NO_XLPFY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPEXT=6 -FGAPEXT=7
-YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : 10 : SwissProt_40 : *
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	66.5	12.4	520	1 APN2 YEAST	P38207 saccharomyc
C 2	66.5	12.4	585	1 HMD1 GOSHI	O64966 gossypium h
C 3	65	12.1	252	1 DHSB SCHPO	P21911 schizosacch
C 4	65	12.8	508	1 HUTH BACSU	P10944 bacillus su
5	64.5	12.7	843	1 PPOC ODOSI	P49467 odontella s
6	64.5	12.7	1002	1 DOR DRONE	Q24314 drosophila
7	63.5	12.5	220	1 Y085 MYCTU	Q10882 mycobacteri
8	63.5	12.5	240	1 PPRH METJA	Q38656 methanococc
C 9	63	11.8	1092	1 LIFR MOUSE	P42703 mus musculu
C 10	62	11.6	259	1 DHSB PARDE	O59662 paracoccu
C 11	62	11.6	489	1 Y225 MYCGE	P47467 mycoplasma
C 12	61.5	11.5	799	1 Y3V2 YEAST	P40897 saccharomyc
C 13	61	12.0	434	1 YNS4 YEAST	Q05131 saccharomyc
14	61	12.0	1076	1 YEM3 YEAST	P40021 saccharomyc
15	61	12.0	2167	1 BM2 YEAST	P39960 saccharomyc
C 16	60.5	11.3	1885	1 RBPQ ACISA	P54891 apple chlor
C 17	59.5	11.7	254	1 PDHR ECOLI	P06957 escherichia
C 18	59.5	11.1	476	1 COXI PLABE	O09252 plasmodium

C 19	59.5	11.1	628	1 HMD2 GOSHI	O64967 gossypium h
20	59.5	11.7	1007	1 CHC2 HUMAN	Q9Y514 homo sapien
C 21	59	11.6	67	1 HMT2 METTH	O27731 methanobact
C 22	59	11.0	75	1 DEF1 CAPAN	Q43413 capsicum an
C 23	59	11.6	276	1 Y009 METJA	Q60320 methanococc
C 24	59	11.0	732	1 YMM1 CAEBL	P34489 caenorhabdi
C 25	59	11.6	870	1 CSX2 SCHPO	Q9UUE2 schizosacch
26	59	11.6	1101	1 DIA2 HUMAN	O60879 homo sapien
C 27	59	11.6	1313	1 VGLM PTPV	P03517 punta toro
C 28	58.5	10.9	169	1 C560 HUMAN	Q99643 h succinate
C 29	58.5	10.9	298	1 DHSB CAEBL	O09545 caenorhabdi
C 30	58.5	10.9	786	1 AAC ACTUT	P29958 actinoplane
C 31	58.5	11.5	2198	1 YLJ2 CAEBL	P34367 caenorhabdi
C 32	58	10.8	265	1 PYRF CANGA	P33283 candida gla
C 33	58	11.4	552	1 VNS1 BTV11	P35932 bluetongue
C 34	58	11.4	552	1 VNS1 BTV17	P14245 bluetongue
C 35	58	10.8	716	1 RRP2 IAZ11	P13175 influenza a
C 36	58	11.4	3054	1 POLG TEV	P04517 t genome po
C 37	58	10.8	5376	1 ZAN MOUSE	O88799 mus musculu
38	57.5	11.3	326	1 OMPL PHOPR	P29739 photobacter
C 39	57.5	11.3	552	1 VNS1 BTV2A	P35931 bluetongue
40	57.5	11.3	575	1 FLA2 CAMJE	P22251 campylobact
41	57.5	11.3	575	1 FLB2 CAMJE	P22252 campylobact
C 42	57.5	10.7	746	1 GYB7 YEAST	P49365 saccharomyc
C 43	57.5	10.7	746	1 GYB7 YEAST	P49365 saccharomyc
C 44	57.5	11.3	1849	1 FSP1 HUMAN	Q9A674 homo sapien
45	57.5	11.3	1727	1 YG44 SCHPO	O60179 schizosacch
				1 ALM1 SCHPO	Q9UTK5 schizosacch

ALIGNMENTS

RESULT 1

ID	APN2_YEAST	STANDARD;	PRT;	520 AA.
AC	P38207:			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DY	13-JUN-2002 (Rel. 41, Last annotation update)			
DE	DNA (apurinic or apyrimidinic site) lyase 2 (EC 4.2.99.18) (AP			
DE	endonuclease 2) (Apurinic-apyrimidinic endonuclease 2).			
GN	APN2 OR ETH1 OR YBL019W OR YBL0443.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
NCBI_TaxID=4932;				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288c;			
RA	Coffeau A., Jonniaux J.-L., Purnelle B., Skala J., de Wergifosse P.,			
RA	van Dyck L.;			
RL	Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.			
RP	CHARACTERIZATION.			
RX	MEDLINE=98438458; PubMed=9765213;			
RA	Johnson R.E., Torres-Ramos C.A., Izumi T., Mitra S., Prakash S.,			
RA	Prakash L.;			
RT	"Identification of APN2, the Saccharomyces cerevisiae homolog of the			
RT	major human AP endonuclease HAP1, and its role in the repair of			
RT	abasic sites.";			
RL	Genes Dev. 12:3137-3143(1998).			
RN	[3]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20357371; PubMed=10806210;			
RA	Unk I., Haracska L., Johnson R.E., Prakash S., Prakash L.;			
RA	"Apurinic endonuclease activity of yeast Apn2 protein.";			
RL	J. Biol. Chem. 275:22427-22434(2000).			
CC	-1- FUNCTION: DNA REPAIR ENZYME THAT HYDROLYZES APURINIC/APYRIMIDINIC			
CC	(AP) SITES AND REMOVES 3'-BLOCKING GROUPS PRESENT AT SINGLE STRAND			
CC	BREAKS OF DAMAGED DNA.			
CC	-1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or			
CC	apyrimidinic site in DNA is broken by a beta-elimination reaction,			
CC	leaving a 3'-terminal unsaturated sugar and a product with a			
CC	terminal 5'-phosphate.			

CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE AP/EXO FAMILY OF DNA REPAIR ENZYMES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z35780; CA84838.1; -
 CC PIR: S45753; S45753.
 CC HSP: P27695; IEN.
 CC SGD: S0000115; APN2.
 CC InterPro: IPR000097; Apendonclsel.
 CC Pfam: PF03372; Exo_endo_phos.
 CC PROSITE: PS00726; AP_NUCLEASE_F1_1; FALSE_NEG.
 CC PROSITE: PS00727; AP_NUCLEASE_F1_2; FALSE_NEG.
 CC PROSITE: PS00728; AP_NUCLEASE_F1_3; 1.
 KW DNA repair; Lyase; Nuclear protein.
 SQ SEQUENCE 520 AA; 59445 MW; E3947C4D904C53FB CRC64;

Alignment Scores:

Pred. No.: 546 Length: 520
 Score: 66.50 Matches: 30
 Percent Similarity: 42.45% Conservative: 15
 Best Local Similarity: 28.30% Mismatches: 37
 Query Match: 12.43% Indels: 24
 DB: 1 Gaps: 4

US-09-856-221-3 (1-291) x APN2_YEAST (1-520)

QY 284 ATGCAGCGCTACTGCTGCTCAATACGATGAGAGATACAGATC----- 234
 DB 247 MetGlyGlyThrLysLeuGluAlaGlnTrpArgAspLysAlaGlnPheIleAsn 266
 QY 233 -----CATGCTGTTGTTTACCGGAGTT----- 210
 DB 267 ProAspThrProHisArgArgIlePheAsnGlnIleLeuAlaAspSerLeuLeuProAsp 286
 QY 209 -----AACATCTTCGTATACAGCGTACTGAAGCTGTCAAAACGAGA----- 168
 DB 287 AlaSerLysArgGlyIleLeuIleAspThrThrArgLeuIleGlnThrArgAsnArgLeu 306
 QY 167 ---TTGTGCCACATGACGGCTTCTGCTTAAGCGCCACTTCATCAGCATCAATTCAGCAAT 111
 DB 307 LysMetThrValTrpAsnMetLeuLys---AsnLeuArgProSerAsnTrpGlySer 325
 QY 110 CATCTTGCCTGCATTTAATACTGTGCAAGCTAATTCATCGCTGTGAGTTTGCGAGCAAT 51
 DB 326 ArgIleAspPheIleLeuVal-SerLeuLysLeuGluArgCysIleLysAlaAlaAspI 345
 QY 50 TCTGCCAGTGGCTCTG 35
 DB 345 eLeuProAspIleLeu 350

RESULT 2

HMD1_GOSHI STANDARD: PRT; 585 AA.
 AC 064966;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 (EC 1.1.1.34) (HMG-
 DE CoA reductase 1).
 GN HMG1
 OS Gossypium hirsutum (Upland cotton).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Malvales; Malvaceae; Gossypium.
 OX NCBI_TaxID=3635;

RN SEQUENCE FROM N.A.
 RP STRAIN:cv. Acala SJ2;
 RA Loquerio L.L.; Wilkins T.A.;
 RT "Two genomic clones encoding 3-hydroxy-3-methylglutaryl-coenzyme A
 CC reductase from cotton (*Gossypium hirsutum* L.).";
 RL (In) Plant Gene Register PGR98-031.
 CC -!- FUNCTION: CATALYZES THE SYNTHESIS OF MEVALONATE. THE SPECIFIC
 CC PRECURSOR OF ALL ISOPRENOID COMPOUNDS PRESENT IN PLANTS.
 CC -!- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) = (S)-3-
 CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
 CC -!- PATHWAY: MEVALONATE BIOSYNTHESIS, PLANT ISOPRENOID BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM. ALSO IN MITOCHONDRIUM AND PLASTID MEMBRANES.
 CC -!- SIMILARITY: BELONGS TO THE HMG-CoA REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF038045; AAC05088.1; -
 CC InterPro: IPR002202; HMG-CoA_red.
 CC InterPro: IPR004554; HMG-CoA_red.
 CC Pfam: PF00368; HMG-coa_red; 1.
 CC PRINTS: PR00071; HMGCOARDTASE.
 CC TIGRFAMs: TIGR00533; HMG_COA_R_NADP; 1.
 CC PROSITE: PS00066; HMG_COA_REDUCTASE_1; 1.
 CC PROSITE: PS00318; HMG_COA_REDUCTASE_2; 1.
 CC PROSITE: PS01192; HMG_COA_REDUCTASE_3; 1.
 CC PROSITE: PS00065; HMG_COA_REDUCTASE_4; 1.
 KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
 KW isoprene biosynthesis; NADP; Multigene family
 FT DOMAIN 1 97 MEMBRANE-BOUND (BY SIMILARITY).
 FT DOMAIN 98 169 LINKER (BY SIMILARITY).
 FT DOMAIN 170 585 CATALYTIC (BY SIMILARITY).
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 77 97 BY SIMILARITY.
 FT ACT_SITE 264 264 BY SIMILARITY.
 FT ACT_SITE 472 472 GENERAL BASE (BY SIMILARITY).
 FT ACT_SITE 570 570 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 574 574 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 585 AA; 62835 MW; 1350E7A9168EBE42 CRC64;

Alignment Scores:

Pred. No.: 551 Length: 585
 Score: 66.50 Matches: 27
 Percent Similarity: 46.34% Conservative: 11
 Best Local Similarity: 32.93% Mismatches: 31
 Query Match: 12.43% Indels: 13
 DB: 1 Gaps: 4

US-09-856-221-3 (1-291) x HMD1_GOSHI (1-585)

QY 274 GTACTGCTGCTCAATACGATGAGAGATACAGATCCATCGCTGT---TTTTCA 218
 DB 164 ValThrValMetThrGluAspGluGluIleLeuArgSerValIcysGlyMetThr 183
 QY 217 CCGGAGTAATCATCTTCTGTCATACAGCGTA-----CTGAAGCTGTCAAAACGAGAT 167
 DB 184 Pro-----SerTyrSerLeuGluSerLysLeuAspCysLysArgAla 198
 QY 166 TGTGCACCATGACGGCTTCTCTTAAGCGCCACTTCATCAGCATCAATTCAGCAATCATC 107
 DB 199 AlalaIleArgGluAlaLeuGlnArgIleThrGlyLysSerLeuSerGlyLeuPro 218
 QY 106 TTGCTCGATTTTAATCTCTGCAAGCTAATTCATCTGCTGCTTGGAGCAATTTCTG 47

Db 219 LeuAspGlyPheAspTyrGluSerIleLeuGlyGlnCysCysGlu-----Met 234
 QY 46 CCAAGT 41
 Db 235 ProVal 236
 RESULT 3
 DMSB_SCHPO STANDARD; PRT; 252 AA.
 AC DMSB_SCHPO
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Succinate dehydrogenase [ubiquinone] iron-sulfur protein,
 DE mitochondrial precursor (EC 1.3.5.1) (IP).
 GN SDH2 OR SPAC140.01
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=2184401; PubMed=11859360;
 RX STRAIN=972;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolke G., Aert R., Robben J., Gymnopoulos B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 [2]
 SEQUENCE OF 89-228 FROM N.A.
 RX MEDLINE=89184541; PubMed=2494655;
 RA Gould S.J., Subramani S., Scheffler I.E.;
 RT "Use of the DNA polymerase chain reaction for homology probing:
 RT isolation of partial cDNA or genomic clones encoding the iron-sulfur
 RT protein of succinate dehydrogenase from several species.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:1934-1938(1989).
 [3]
 RP ERRATUM.
 RA Gould S.J., Subramani S., Scheffler I.E.;
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2556-2556(1993).
 CC -1- CATALYTIC ACTIVITY: Succinate + ubiquinone = fumarate + ubiquinol.
 CC -1- COFACTOR: BINDS THREE DIFFERENT IRON-SULFUR CLUSTERS: A 2FE-2S,
 CC -1- PATHWAY: Tricarboxylic acid cycle.
 CC -1- SUBUNIT: COMPOSED OF A 27 kDa IRON PROTEIN (IP) AND A 70 kDa
 CC FLAVOPROTEIN (FP).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL163191; CAB86412.1; -.
 DR PIR: E32394; E32394.
 DR HSP: P00364; IFUM.
 DR InterPro: IPR000564; 2FE2S-ferredoxin.
 DR InterPro: IPR001450; 4FE4S-ferredoxin.
 DR InterPro: IPR004489; DhsB.
 DR InterPro: IPR001041; Ferredoxin.
 DR Pfam: PF00111; fer2; 1.
 DR TIGRfam: TIGR00384; dhsb; 1.
 DR PROSITE: PS00197; 2FE2S-FERREDOXIN; 1.
 DR PROSITE: PS00198; 4FE4S-FERREDOXIN; 1.
 KW Tricarboxylic acid cycle; Iron-sulfur; Oxidoreductase; Mitochondrion;
 KW Electron transport; 3Fe-4S; 4Fe-4S; Transit peptide.
 FT TRANSIT 1 ? MITOCHONDRION.
 FT CHAIN ? 252 SUCCINATE DEHYDROGENASE [UBIQUINONE]
 FT METAL 75 75 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 80 80 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 83 83 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 95 95 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 165 165 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 168 168 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 171 171 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 175 175 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
 FT METAL 222 222 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
 FT METAL 228 228 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
 FT METAL 232 232 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 252 AA; 28595 MW; 87A3DC0B3742DEEL CRC64;
 Alignment Scores:
 Pred. No.: 7.83 Length: 252
 Score: 65.00 Matches: 22
 Percent Similarity: 46.77% Conservative: 7
 Best Local Similarity: 35.48% Mismatches: 16
 Query Match: 12.15% Indels: 17
 DB: 1 Gaps: 2
 US-09-856-221-3 (1-291) x DMSB_SCHPO (1-252)
 QY 258 ATACCATCAAGAGATACAGATCCATCGCTGTTTTCACCGAGTAAACATCTTCGT 199
 Db 59 ILeIySIleIySAsnGluInAspProThrLeuThrPheArg----- 73
 QY 198 CATACAGCGTACTGAAGCTGTCAAACGAGATGTGCACCATGCGCTTTCCTTAAGCG 139
 Db 74 -----SerCysArgGluGlyIle-CysGlySerCysAla----- 84
 QY 138 CCATTCATCAGCATCAATTTTCAGCAATCATCTTCCTGCAATTTTAATCTGCAAGCG 79
 Db 85 -----MetAsnIleAsnGlySerAsnThrLeuAlaCysIleCysAsnIleLysLysAs 102
 QY 78 TAAT 75
 Db 102 pasn 103
 RESULT 4
 HUTH_BACSU STANDARD; PRT; 508 AA.
 AC P10944;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
 GN HUTH
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;


```

CC -----
DR EMBL; Z67753; CAA91745.1; -.
DR HSP; O9KWU6; LHOM.
DR InterPro; IPR000722; RNA_pol_A.
DR Pfam; PF00523; RNA_pol_A; 1.
KW transferase; transcription; DNA-directed RNA polymerase; Chloroplast.
SQ SEQUENCE 843 AA; 98521 MW; DF31C1527EC5E58E CRC64;

Alignment Scores:
Pred. No.: 9.97 Length: 843
Score: 64.50 Matches: 18
Percent Similarity: 46.67% Conservative: 24
Best Local Similarity: 20.00% Mismatches: 45
Query Match: 12.72% Indels: 3
DB: 1 Gaps: 2

US-09-856-221-3 (1-291) x RPOC_ODOSI (1-843)
QY 22 GAAGCGAGGATGCGAGAGCACTGGCAGAAATGCTGCAAACTCAAGCAGCTGAATTAGCT 81
Db 391 GluLeuGluAspLeuAspIleIleLysGluIleGlnArgThrArgLysPheIleValIle 410
QY 82 TTGCAGAGTATTAAATATGACGAGCAAGATGATGCTGAAATGTAGT-----CTGATGAAG 135
Db 411 CysSerLysIleLeuHisLysGluLysProIleTyrHisPheLeuValGtTTPheArgLys 430
QY 136 TGGCGCTTGAAGNAAGCGCTATGCTGCACATCTGTTTCGACAGCTTCAGTACGCTGT 195
Db 431 TrpLeuLeuGlnArgAlaTyrLysLeuArgAspGlnAlaIle---LysArgIleArgIle 449
QY 196 ATGACGAAGATGTTAACTCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 255
Db 450 LeuGluAsnLeuLeuAlaThrGlySerAsnProAlaTrpMetIleLeuThrIleLeuPro 469
QY 256 TATTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 285
Db 470 ValIleProAlaLeuArgProMetIle 479

RESULT 6
DOR_DROME
ID DOR_DROME STANDARD; PRT; 1002 AA.
AC Q24314;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Deep orange protein.
GN DOR OR EG171E4.1
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97218037; PubMed=9065698;
RA Shestopal S.A., Makum I.V., Belyaeva E.S., Ashburner M.;
RT "Molecular characterization of the deep orange (dor) gene of
RL Drosophila melanogaster."
RL Mol. Gen. Genet. 253:642-648(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Gallibert F., Borkova D.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikotli F.,
RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,
RA Glover D.M.;

```

```

RT -----
RT melanogaster";
RL Science 287:2220-2222(2000).
CC -1- SIMILARITY: SOME, TO YEAST PEP3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL; X86683; CAA60382.1; -.
DR EMBL; AL021726; CAA16809.1; -.
DR Flybase; FBgn0000482; dor.
DR InterPro; IPR000547; Clathrin_repeat.
DR SMART; SM00299; CLH; 1.
DR SMART; SM00184; RING; 1.
KW Zinc-finger; Transmembrane.
FT ZN_FING 885 910 C3H2C-TYPE.
FT TRANSMEM 971 991 POTENTIAL.
FT CONFLICT 169 169 A -> P (IN REF. 1).
FT CONFLICT 581 581 O -> H (IN REF. 1).
FT CONFLICT 865 865 A -> V (IN REF. 1).
SQ SEQUENCE 1002 AA; 115305 MW; D59690A0FC95182F CRC64;

Alignment Scores:
Pred. No.: 10.1 Length: 1002
Score: 64.50 Matches: 22
Percent Similarity: 42.16% Conservative: 21
Best Local Similarity: 21.57% Mismatches: 36
Query Match: 12.72% Indels: 23
DB: 1 Gaps: 3

US-09-856-221-3 (1-291) x DOR_DROME (1-1002)
QY 1 AATACCTTCTCAACATTACTGAACGGCAGATGCGAGACGACTGGCAGAAATGCTGCAA 60
Db 698 AsnPheLeuLeuHisLeuTyrAlaGluHisGluProLysLeuLeuMetLysTyrLeu 717
QY 61 ACTCAAGCGAGTGAATTAGCTTTCGACAGTATTAAATG-----99
Db 718 IleGlnGlyArgAspLysLeuValHisTyrAspIleTyrTyrAlaHisLysValCys 737
QY 100 -----CAGGCAAGATGATCTCTGCAAAATGATGCTGATGCTGATGCTGCTG 141
Db 738 ThrAspLeuAspVallyGluAlaArgValPheLeuGluCysMetLeuArgLysTrp--- 756
QY 142 TTAAGGAAAGCGTCATGTCACAAATCTCGTTTTCACAGCTTCAGTACGCTGATGAGC 201
Db 757 -----IleSerAlaValAspLeuAlaLeuThrPheAspMetLysLeuAlaLys 772
QY 202 AAGATGTTA-----ACTCGGTGAAACAAACAGCGATGCTGATCTCTCT 246
Db 773 GluThrAlaSerArgProSerAspSerLysIleArgLysLeuTrpLeuArgIleAla 792
QY 247 CTCAT 252
Db 793 TyrHis 794

RESULT 7
Y085_MYCTU
ID Y085_MYCTU STANDARD; PRT; 220 AA.
AC Q10882;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Y0085.
GN Y0085 OR MY0092 OR MTCY251.03.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

```


QY 99 GCAGGCAAGATGCTGCTGAATGATGCTGATGAGTGGCGCTTAAGAA---AGCCGT 155
 D 149 AlaAspLeuValIleGlyThrAsnValAspGlyValTyAspLysAspProAsnLys 168
 QY 156 CATGGTGCACATCTCTGTTTTCAGCAGCTTCACGTACGCTGATGACGAGATGTTAACTCC 215
 D 169 TyrGluAspAlaLysLysPheAspLysMetSerAla----- 180
 QY 216 GGTGAAACACAGCATGATCTATCTCTTCATCGGTATGACGACGACGATGAG 275
 D 181 -----LysGluLeuValAspLeuAlaIleSerSerSerLeuLysAlaGlySerSerSer 198
 QY 276 GCCCTGCATATGGT 290
 D 199 ValValAspLeuLeu 203
 RESULT 9
 ID LIFR_MOUSE STANDARD; PRT; 1092 AA.
 AC P42703;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Leukemia inhibitory factor receptor precursor (LIF-R) (D-factor/LIF
 receptor).
 GN LIFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 MM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=92007727; PubMed=1915266;
 RA Gearing D.P., Thut C.J., Vandenbos T., Gimpel S.D., Delaney P.B.,
 RA King J., Price V., Cosman D., Beckmann M.P.,
 RT "Leukemia inhibitory factor receptor is structurally related to the
 RT IL-6 signal transducer, gp130.";
 RL EMBO J. 10:2839-2848(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX STRAIN=ICR; TISSUE=Liver;
 RC MEDLINE=94039833; PubMed=7901054;
 RA Tomida M., Yamamoto-Yamaguchi Y., Hozumi M.;
 RT "Pregnancy associated increase in mRNA for soluble D-factor/LIF
 RT receptor in mouse liver.";
 RL PERS Lett. 334:193-197(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=94334302; PubMed=8056772;
 RA Tomida M., Yamamoto-Yamaguchi Y., Hozumi M.;
 RT "Three different cDNAs encoding mouse D-factor/LIF receptor.";
 RL J. Biochem. 115:557-562(1994).
 CC -|- FUNCTION: SIGNAL-TRANSDUCING MOLECULE. MAY HAVE A COMMON PATHWAY
 CC WITH GPI30. THE SOLUBLE FORM INHIBITS THE BIOLOGICAL ACTIVITY OF
 CC LIF BY BLOCKING ITS BINDING TO RECEPTORS ON TARGET CELLS.
 CC -|- SUBUNIT: HETERODIMER COMPOSED OF LIFR AND GPI30.
 CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THERE IS A
 CC MEMBRANE-BOUND AND A SECRETED FORM.
 CC -|- ALTERNATIVE PRODUCTS: 2 isoforms: 1/membrane (shown here) and
 CC 2/secreted; may be produced by alternative splicing.
 CC -|- TISSUE SPECIFICITY: PLACENTA, LIVER, KIDNEY, HEART, LUNG, BRAIN,
 CC AND EMBRYOS. THE LIVER MAY BE THE PRIMARY SITE OF SYNTHESIS OF
 CC THE SECRETED FORM.
 CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -|- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -|- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC
 DR ENBL; S73496; AAC60698.1; -
 DR ENBL; S73495; AAC60697.1; -
 DR ENBL; D26177; BAA05165.1; -
 DR ENBL; D17444; BAA04258.1; -
 DR MGD; MGI:96788; Lifr.
 DR InterPro; IPR002996; CR1A.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003529; Hemtopoptn_L_F2.
 DR Pfam; PF00041; fn3; 4
 DR SMART; SM00060; FN3; 3.
 DR PROSITE; PS01353; HEMATOPO_REC_L_F2; 1.
 DR Receptor; Transmembrane; Glycoprotein; Immunoglobulin domain; Signal;
 KW Alternative splicing; Repeat.
 FT SIGNAL 1 43 POTENTIAL.
 FT CHAIN 44 1092 LEUKEMIA INHIBITORY FACTOR RECEPTOR.
 FT DOMAIN 44 828 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 829 853 POTENTIAL.
 FT DOMAIN 854 1092 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 53 63 BY SIMILARITY.
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 647 647 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 724 724 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 718 719 AP -> EA (IN ISOFORM 2).
 FT VARSPLIC 720 1092 MISSING (IN ISOFORM 2).
 SQ SEQUENCE 1092 AA; 122573 MW; 6F02B8C8E154DE70 CRC64;
 Alignment Scores:
 Pred. No.: 15.5 Length: 1092
 Score: 63.00 Matches: 27
 Percent Similarity: 38.78% Conservative: 11
 Best Local Similarity: 27.55% Mismatches: 30
 Query Match: 11.78% Indels: 30
 DB: 1 Gaps: 5
 US-09-856-221-3 (1-291) x LIFR_MOUSE (1-1092)
 QY 290 CACCATATCAGGCGCGTACT-----GCTGCTCAATACCATGATGAAGA 246
 D 378 HIsArgIleGluGlyLeuThrAsnGluThrTyArgLeuGlyValGlnMetHisProGly 397
 QY 245 GAGATACAGATCCATCGCTGTTTTCACCGGAGTAAAC----- 207
 D 398 Gln---GluIleHisAsnPheThrLeuThrGlyArgAsnProLeuGlyGlnAlaGlnSer 416
 QY 206 ATCTGCTATACAGCGTACTGAAGCTGTC----- 177
 D 417 AlaValIleAsnValThrGluArgValAlaProHisAspProThrSerLeuLysVal 436
 QY 176 -----AAACAGAGATTGTGCACGATGACGGCTTTCCTTAAGCGCCACTTCATCAGC 126
 D 437 LysAspIleAsnSerThrValValThrPheSerTrpTyLeuProGlyAsnPheThrLys 456
 QY 125 ATCAATTTCAGCAATCATCTTCCTGCATTTTAACTACTCTGCAAGCTAATTC 72
 D 457 IleAsn-----LeuLeuCysGlnIleGluIleCysLysAlaAsnSer 470
 RESULT 10
 DHSB-PARDE

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; Z35159; CNA84524.1; -;
DR EMBL; L33832; AAA57132.1; -;
DR EMBL; U18917; AAB64682.1; -;
DR SGD; S0000957; BEM2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGEF_CDC25.
DR InterPro; IPR000198; RhogAP.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00620; RhogAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
DR SMART; SM00324; RhogAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW GTPase activation.
FT DOMAIN 1914..1948 PH.
FT DOMAIN 1981..2134 RHO-GAP.
FT DOMAIN 16..25 POLY-SER.
FT DOMAIN 35..43 POLY-SER.
FT DOMAIN 58..63 POLY-HIS.
FT DOMAIN 198..208 POLY-ASN.
FT DOMAIN 253..260 POLY-SER.
FT DOMAIN 1161..1165 POLY-THR.
SQ SEQUENCE 2167 AA; 245428 MW; DB7F4CD417E898F0 CRC64;

Alignment Scores:

Pred. No.: 28.8 Length: 2167
Score: 61.00 Matches: 21
Percent Similarity: 49.35% Conservative: 17
Best Local Similarity: 27.27% Mismatches: 21
Query Match: 12.03% Indels: 18
DB: 1 Gaps: 3

US-09-856-221-3 (1-291) x BEM2_YEAST (1-2167)

QY 31 GATCGACAGCACTGGCAGAAATTCCTGCAACTCAAGCG-----AGTGAATTA----- 78
Db 1709 GluSerGluAlaIleSerGluArgPheGlnGluGlnGlyValPheAsnGluIleLeuVal 1728
QY 79 ---GCTTTCAGAGTATTAAATCAGCAGCAAGATGATTGCTGAAATTGATGCTGATGAAG 135
Db 1729 AsnGluIleGluLysIleLysArgGluAlaArgLysLeuGluValLeuLeuAspGlnGlu 1748
QY 136 TGGCGCTTAAGGAAGCGTCATGGTCACAATCTCGTTTGTGACAGCTTCAGTACGCTGT 195
Db 1749 LysIleLeuLysAsnSerAlaAlaLeuHisGlnAlaVal----- 1761
QY 196 ATCAGCAAGATTTAACTCCGGTCAAAACACAGCGATGATCTATCTCT 246
Db 1762 -----ProLysLysAsnArgLysSerValIleIleSer 1772

Search completed: January 15, 2003, 15:46:31
Job time : 10.8834 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 15, 2003, 15:44:13 ; Search time 27.4148 Seconds
(without alignments)
4374.264 Million cell updates/sec

Title: US-09-856-221-3
Perfect score: 507
Sequence: 1 aatacctgtcacaattac.....tacggccgtcatatggtgc 291

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/uspto.spool/us09856221/runat_15012003_153925_12457/app_query.fasta.1.1948
-DB-SPTREMBL_21 -QPM=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR.SCORE=pcpt -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221@cgn_1_1_171_runat_15012003_153925_12457 -NCPU=3
-NO.XLPXY -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*
17: sp.archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	399	78.7	2523	2 Q93RP0	Q93rp0 xenorhabdus

2	293	57.8	2376	2	Q9F923
3	220	43.4	2516	2	Q9RN43
4	208	41.0	2504	2	O85160
5	184	36.3	2538	2	Q93RN7
6	109	21.5	1565	2	O85156
7	71	14.0	1693	2	Q99Y37
8	71	14.0	1693	2	Q9F5F7
9	69	13.6	1379	10	Q07893
10	68	13.4	299	5	O16734
11	68	13.4	329	16	O34803
12	68	13.4	626	3	O59966
13	68	13.4	1383	10	O65808
14	68	13.4	1693	2	Q986L2
15	68	13.4	1693	16	Q80634
16	67	13.2	1189	2	O85152
c 17	67	12.5	1345	5	O96236
18	67	13.2	1380	10	Q40001
19	66	13.0	218	16	Q927Y1
20	66	12.3	1349	12	Q9DWB6
c 21	66	13.0	1463	5	Q9GVZ0
c 22	65.5	12.2	930	10	O22161
23	65	12.8	138	2	Q9X2K5
24	65	12.8	274	16	Q97N99
25	65	12.8	1279	2	Q93RP4
26	65	12.8	1382	10	O22435
c 27	64.5	12.1	228	8	O8WCM5
c 28	64.5	12.1	370	12	Q81967
29	64.5	12.7	1002	5	Q95R67
30	64.5	12.7	1002	5	Q9W570
31	64.5	12.7	1033	16	Q8YNX6
c 32	64.5	12.1	1036	5	Q76974
c 33	64.5	12.1	1037	5	Q9N9D8
c 34	64.5	12.1	1127	5	Q27493
35	64	12.6	464	5	Q9V8D8
36	64	12.0	782	5	O61073
37	64	12.6	784	10	Q9XE10
38	64	12.6	1040	11	Q92518
39	64	12.6	1561	11	O88286
40	63.5	12.5	475	16	Q8XYA8
c 41	63.5	11.9	578	10	Q9FM59
c 42	63	12.4	186	10	O22590
43	63	12.4	256	16	Q9AB37
c 44	63	11.8	338	5	Q9VW96
45	63	12.4	350	16	Q9ZJF8

ALIGNMENTS

RESULT 1
Q93RP0 PRELIMINARY: PRT: 2523 AA.
ID Q93RP0
AC Q93RP0
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Xptal protein.
GN XPTAL.
OS Xenorhabdus nematophilus.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Xenorhabdus.
OX NCBI_TaxID=628;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMF1296;
RX MEDLINE=2121813; PubMed=11319082;
RA Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus
pmf1296.";
RL Appl. Environ. Microbiol. 67:2062-2069(2001).
DR EMBL; AJ308438; CAC38401.1;
DR InterPro; IPR000953; Chromo.
DR SMART; SM00298; CHROMO; 1.
SQ SEQUENCE 2523 AA; 286999 MW; 3159852E055B5B1 CRC64;

Alignment Scores:
 Pred. No.: 1,01e-41 Length: 2523
 Score: 399.00 Matches: 85
 Percent Similarity: 97.92% Conservative: 9
 Best Local Similarity: 88.54% Mismatches: 2
 Query Match: 78.70% Indels: 1
 DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x Q93RP0 (1-2523)

QY 1 AATACCTTGCTCAACATTACTGAAGCGCAGGATGCAGAACACTGGCAGAAATGCTGCAA 60
 Db 2054 AsnThrLeuLeuSerIleThrGluArgGlnAspAlaGluAlaLeuAlaGluLeuGln 2073

QY 61 ACTCAAGGCAGTGAATTACCTTTCAGAGTATTAAATGCAGGC-AAGATGATTGCTGAA 119
 Db 2074 ThrGlnGlySerGluLeuAlaLeuGlnSerIleLeuMetGlnAspValMetAlaGlu 2093

QY 120 ATTGATGCTGATGAAGTGGCGTTAAGGAAAGCGCGTCATGGTGCAACATCTCGTTTGGAC 179
 Db 2094 IleAspAlaAspIleLeuAlaLeuGlnSerArgHisGlyAlaGlnSerArgPheAsp 2113

QY 180 AGCTTCAGTACGCTGATCAGCAAGATGTTAACTCCGGTGAAACAAAGCGATGGATCTG 239
 Db 2114 SerPheAsnThrLeuTyrraspGluAspValAsnAlaGlyGluLeuGlnAlaMetAspLeu 2133

QY 240 TATCTCTCTTCATCGGTATTGAGCACCAGCAGTACGGCCCTGCATATG 287
 Db 2134 TyrLeuSerSerValLeuSerThrSerGlyThrAlaLeuHisMet 2149

RESULT 2
 Q9F923 ID Q9F923 PRELIMINARY: PRT; 2376 AA.
 AC Q9F923;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Sepa.
 GN Serratia entomophila.
 OS Serratia entomophila.
 OG Plasmid PADAP.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 OX NCBI_TaxID=42906;
 RN [1]
 RC STRAIN=AIW02;
 RX MEDLINE=20416224; PubMed=10960097;
 RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
 RT "Plasmid-located pathogenicity determinants of Serratia entomophila,
 RT the causal agent of amber disease of grass grub, show similarity to
 RT the insecticidal toxins of Photobacterium luminescens.";
 RL J. Bacteriol. 182:5127-5138(2000).
 DR EMBL; AF135182; AAG09642.1; -;
 KW Plasmid.
 SQ SEQUENCE 2376 AA; 262649 MW; E418DAC22DBB7BF CRC64;

Alignment Scores:
 Pred. No.: 3e-28 Length: 2376
 Score: 293.00 Matches: 65
 Percent Similarity: 80.21% Conservative: 12
 Best Local Similarity: 67.71% Mismatches: 19
 Query Match: 57.79% Indels: 1
 DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x Q9F923 (1-2376)

QY 1 AATACCTTGCTCAACATTACTGAAGCGCAGGATGCAGAACACTGGCAGAAATGCTGCAA 60
 Db 1907 AsnThrLeuLeuGlyIleThrGluArgGlnAspAlaGluAlaLeuAlaGluLeuGln 1926

QY 61 ACTCAAGGCAGTGAATTACCTTTCAGAGTATTAAATGCAGGC-AAGATGATTGCTGAA 119

Db 1927 ThrGlnGlySerGluLeuIleArgGlnGlnLeuArgGlnGlnAspAsnValLeuGluGlu 1946
 QY 120 ATTGATGCTGATGAAGTGGCGTTAAGGAAAGCGCGTCATGGTGCAACATCTCGTTTGGAC 179
 Db 1947 IleAspAlaAspIleAlaLeuGluSerArgArgGlyAlaGlnMetArgPheGlu 1966
 QY 180 AGCTTCAGTACGCTGATCAGCAAGATGTTAACTCCGGTGAAACAAAGCGATGGATCTG 239
 Db 1967 ArgTyrrLysValLeuTyrrGluAlaAspValAsnThrGlyGluLysGlnAlaMetAspLeu 1986

QY 240 TATCTCTCTTCATCGGTATTGAGCACCAGCAGTACGGCCCTGCATATG 287
 Db 1987 TyrLeuSerSerValLeuSerAlaSerThrAlaAlaLeuPheLeu 2002

RESULT 3
 Q9RN43 ID Q9RN43 PRELIMINARY: PRT; 2516 AA.
 AC Q9RN43;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Toxin A (Toxin complex protein).
 GN TCDA.
 OS Photobacterium luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Photobacterium.
 OX NCBI_TaxID=29488;
 RN [1]
 RC STRAIN=W-14;
 RA Merlo D.J., Wegrich L.M., Roberts J.L., Petell J.K.;
 RT "Expression of Photobacterium luminescens tcda gene in maize confers
 RT corn rootworm tolerance.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC STRAIN=W14;
 RX MEDLINE=21185117; PubMed=11286884;
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
 RA French-Constant R.H.;
 RT "The tc genes of Photobacterium: a growing family.";
 RL Trends Microbiol. 9:185-191(2001).
 DR EMBL; AF188483; AAF05542.1; -;
 DR EMBL; AF346500; AALJ8486.1; -;
 DR InterPro: IPR001589; Actbind.actnin.
 DR PROSITE: PS00019; ACTININ.1; UNKNOWN.1
 SQ SEQUENCE 2516 AA; 282952 MW; 8250A0650B614B99 CRC64;

Alignment Scores:
 Pred. No.: 5.74e-19 Length: 2516
 Score: 220.00 Matches: 50
 Percent Similarity: 74.19% Conservative: 19
 Best Local Similarity: 53.76% Mismatches: 24
 Query Match: 43.39% Indels: 1
 DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x Q9RN43 (1-2516)

QY 1 AATACCTTGCTCAACATTACTGAAGCGCAGGATGCAGAACACTGGCAGAAATGCTGCAA 60
 Db 2038 SerThrLeuGlnAsnIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuLeuGln 2057

QY 61 ACTCAAGGCAGTGAATTACCTTTCAGAGTATTAAATGCAGGC-AAGATGATTGCTGAA 119
 Db 2058 AsnGlnAlaAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 2077

QY 120 ATTGATGCTGATGAAGTGGCGTTAAGGAAAGCGCGTCATGGTGCAACATCTCGTTTGGAC 179
 Db 2078 LeuAspAlaGluLysThrValLeuGluLysSerIleValGlnSerArgPheAsp 2097

QY 180 AGCTTCAGTACGCTGATCAGCAAGATGTTAACTCCGGTGAAACAAAGCGATGGATCTG 239
 Db 180 AGCTTCAGTACGCTGATCAGCAAGATGTTAACTCCGGTGAAACAAAGCGATGGATCTG 239

RN SEQUENCE FROM N.A.
 RP STRAIN-W14;
 RX MEDLINE=21195117; PubMed=11286884;
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
 RA French-Constant R.H.;
 RT "The to genes of Photobacterium: a growing family.";
 RL Trends Microbiol. 9:185-191(2001).
 DR EMBL; AF047028; AAC38629.1; -;
 DR EMBL; AF346499; AAL18472.1; -;
 SQ SEQUENCE 1565 AA; 175717 MW; 66C3AB96C5FA8397 CRC64;

Alignment Scores:
 Pred. No.: 7,046-05 Length: 1565
 Score: 109.00 Matches: 31
 Percent Similarity: 56.52% Conservative: 21
 Best Local Similarity: 33.70% Mismatches: 40
 Query Match: Indels: 1
 DB: Gaps: 2

US-09-856-221-3 (1-291) x 085156 (1-1565)
 QY 7 TTGCTCAACATTACTGAACGCGAGATGCGAGCAGCTGGCAGAAATTCCTGCAACTCAA 66
 Db 1076 LeuLeuSerLeuLeuGluArgSerGluArgAlaCysGlnGluGluLeuAlaGlnGln 1095
 QY 67 GCGAGTGAATAGCTTTGCGAGAGTATTAAATGACGCGC-AGATGATGCTGGAATTTGAT 125
 Db 1096 LeuLeuAspMetSerTyrAlaIleThrLeuGlnGlnGlnAlaLeuAspGlyLeuAla 1115
 QY 126 GCTGATGAAGTGGCGCTTAAGGAAACGCTCATGCTGCACAAATCTCTGTTTTCACAGCTTC 185
 Db 1116 AlaAspArgLeuAlaLeuAlaSerGlnAlaThrAlaGlnGlnArgHisAspHisTyr 1135
 QY 186 AGTACGCTGTATGACGAGATGTTAACTCCGGTGAAACACAGCGATGATCTATCTC 245
 Db 1136 TyrThrLeuTyrGlnAsnAspSerSerAlaGluGlnLeuValMetAspThrGlnThr 1155
 QY 246 TCTTCATCGTATTGACGACGACGATGACGCGCTG 281
 Db 1156 SerAlaGlnSerLeuIleSerSerThrGlyVal 1167

RESULT 7
 Q99Y37 PRELIMINARY; PRT; 296 AA.
 AC Q99Y37
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Similar to several eukaryotic hypersensitive-induced response
 DE proteins.
 GN SPX1884.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL; AE006613; AAK34596.1; -;
 DR InterPro; IPR001107; Band7.
 DR Pfam; PF01145; Band7; 1.
 DR SMART; SM00244; PHB; 1.
 KW Complete proteome.
 SQ SEQUENCE 296 AA; 33066 MW; 530DA0BFB7E7340 CRC64;

Alignment Scores:
 Pred. No.: 4,02 Length: 296
 Score: 71.00 Matches: 25
 Percent Similarity: 46.67% Conservative: 17
 Best Local Similarity: 27.78% Mismatches: 38
 Query Match: Indels: 10
 DB: Gaps: 2

US-09-856-221-3 (1-291) x Q99Y37 (1-296)
 QY 36 AGAAGCAGCTGGCAGAAATGCTGCAAACTCAAGCAGTGAATTAGCTTTGCAGAGTATTAA 95
 Db 175 GlnSerMetAsnGluIleAsnAlaGlnArgLysArgValAlaGlnGluLeuAla 194
 QY 96 AATCAGCGCAAGATGATGCTGCAAAATGATGCTGATGAAGTGGCGCTTAAGAAAGCCGT 155
 Db 195 AsnAlaAspLysIleLysIleValThrAlaAlaGluAlaGluLysAspArgLeu 214
 QY 156 CATGGT-----CCCAATCTCGT-----TTTGACAGCTTC 185
 Db 215 HisGlyValGlyIleAlaGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 234
 QY 186 AGTACGCTGTATGACGAGATGTTAACTCCGGTGAAACACAGCGATGATCTATCTC 245
 Db 235 GlnGluLeuLysGluAlaAsnIleSerLeuAsnGluGlnIleMetSerIleLeuLeu 254
 QY 246 TCTTCATCGTATTGACGACGACGATGACG 275
 Db 255 ThrAsnGlnTyrLeuAspThrLeuAsnThr 264

RESULT 8
 Q9F5F7 PRELIMINARY; PRT; 1693 AA.
 AC Q9F5F7
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Riorf93 protein.
 GN RIORF93.
 OS Agrobacterium rhizogenes.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724;
 RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
 RA Yoshida K.;
 RT "The complete nucleotide sequence of a Ri (root inducing) plasmid
 RT indicates its chimerical structure between Ti and Sym plasmids.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724;
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
 RT "Analysis of unique variable region of a plant root inducing plasmid,
 RT pRi1724, by the construction of its physical map and library.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724;
 RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
 RT "Genome structure of Ri plasmid (1): Construction of linking library
 RT and physical map of pRi1724 in Japanese Agrobacterium.";
 RL Nucleic Acids Symp. Ser. 39:189-190(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724;
 RX MEDLINE=20241294; PubMed=10780382;
 RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
 RA Yoshida K.;
 RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and

RT its flanking regions of pRI1724 in Japanese *Agrobacterium*
rhizogenes.⁺;
RL Nucleic Acids Symp. Ser. 42:67-68(1999).
DR EMBL: AP02086; BAB16212.1; -;
DR HSP: P14385; 2ADM.

DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002296; N12N6_mtfase.
DR InterPro: IPR002052; N6_Mtase.
DR InterPro: IPR000051; SAM_bind.
DR Pfam: PF00271; helicase_C; 1.
DR PRINTS: PR00507; N12N6MTFRASE.
DR SMART: SM00487; DEXdc; 1.
DR PROSITE: PS00092; N6_MTASE; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 1693 AA; 188083 MW; 62B8E7E786782555 CRC64;

Alignment Scores:
Pred. No.: 4.8 Length: 1693
Score: 71.00 Matches: 21
Percent Similarity: 54.10% Conservative: 12
Best Local Similarity: 34.43% Mismatches: 26
Query Match: 14.00% Indels: 2
DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x 09F5F7 (1-1693)

QY 111 ATGCTGAATTCATGCTGATGAAGCGCGCTTAAGGAAGCGCTCATGTCGACAACTCT 170
DB 1033 ILEALAGLIEGLYVALASPGINILFIEVALASPGIUALAGLUPHEARGLYSLEU 1052
QY 171 CGTTTT---GACAGCTTCAGTACGCTGTATGACGAGAGATGTAACCCGCGTGAACAA 227
DB 1053 SerPheAlaThrAsnMetSerThrLeuLysGlyIleAspProAsn---GlySerGlnArg 1071
QY 228 GCGATGATCTGTACTCTCTTCATCGTATTCAGCACCAGCAGTACGGCCCTGCATATG 287
DB 1072 AlatrAspLeuTyValLysSerArgPheLeuGluThrLysAsnProGlyArgAlaLeu 1091
QY 288 GTG 290
DB 1092 Val 1092

RESULT 9
Q07893
ID Q07893 PRELIMINARY; PRT; 1379 AA.
AC Q07893;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Protoporphyrin IX:Mg chelatase.
GN OLIVE.
OS Antirrhinum majus (Garden snapdragon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
OX NCBI_TaxID=4151;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILD TYPE, AND CV. J.I.98;
RX MEDLINE=94008977; PubMed=8404842;
RA Hudson A., Carpenter R., Doyle S., Coen E.S.;
RT "Olive: a key gene required for chlorophyll biosynthesis in
Antirrhinum majus."
RL EMBL J. 12:3711-3719(1993).
DR EMBL: X73144; CAA51664.1; -;
DR InterPro: IPR003672; CbnN/Mg_chitase.
DR Pfam: PF02514; cbnN-Mg_chel; 1.
SQ SEQUENCE 1379 AA; 133217 MW; 0088E819984F505D CRC64;

Alignment Scores:
Pred. No.: 8.45 Length: 1379
Score: 59.00 Matches: 24

Percent Similarity: 40.96% Conservative: 10
Best Local Similarity: 28.92% Mismatches: 25
Query Match: 13.61% Indels: 24
DB: 10 Gaps: 3

US-09-856-221-3 (1-291) x 007893 (1-1379)

QY 99 GCAGCGAAGATGATGCTGAATGAT----- 125
DB 1102 AlaValLysMetValAlaGluLeuAspGluProValGluGlnAsnPheValArgLysHis 1121
QY 126 -----GCTGATGAAGTGGCGCTTAAGGAAGCGCTCATGTCGACAACTCTCGT 173
DB 1122 AlaLeuGluGlnAlaLysGluLeuGlyVal---GluValArgGluAlaAlaSerArgIle 1140
QY 174 TTTCACAGCTTCAGTACGCTGTATGACGAGATGTTAAC----- 212
DB 1141 PheSerAsnAlaSerGlySerTyrSerSerAsnIleAsnLeuAlaValGluAsnSerSer 1160
QY 213 ---TCCGGTGAACAAACAGCGATGATCTCTCTCTCTCTCTCGTATTGACGACGACG 269
DB 1161 TrpAsnAspGluLysGlnLeuGlnAspMetTyrLeuSerArgLysSerPheAlaPheAsp 1180
QY 270 AGTACGGCC 278
DB 1181 SerAspAla 1183

RESULT 10
O16734
ID O16734 PRELIMINARY; PRT; 299 AA.
AC O16734;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE K09F6.7 protein.
GN K09F6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilsson R., Alnscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighthning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roope A., Saunders D., Showkeen R.,
RA Saldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Scheet P.;
RT "The sequence of C. elegans cosmid K09F6.";
RL submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: AF016683; AAB66197.1; -;
DR HSSP: P15919; 1RMD.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; zf-C3HC4; 1.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 15, 2003, 15:39:33 ; Search time 30.7357 Seconds
(without alignments)
2878.683 Million cell updates/sec
Title: US-09-856-221-4
Perfect score: 605
Sequence: 1 gtgaagcgcgatttgcaa.....atttgaaaaagatgaacgg 332
Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=framet.n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09856221/runat_15012003_153924_12438/app_query_fasta_1.1948
-DB=A_Geneseq_101002_QPWT-fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=-1 -MATRIX=DIOSUMC2 -TRANS-human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221_@cgn_1_1_60_@runat_15012003_153924_12438 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NRG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	446	73.7	579	18	AAW18301	Photorhabdus lumin
2	446	73.7	579	19	AAW56574	Toxin TcdAIII, enc
3	446	73.7	2516	18	AAW17899	Photorhabdus lumin
4	446	73.7	2516	19	AAW56572	Toxin TcdA, encode
5	446	73.7	2516	22	AAW72609	Photorhabdus tcdA
6	446	73.7	2517	22	AAW72611	Modified Photorhab
7	446	73.7	2537	22	AAW72614	TcdA toxin-zeln ER
8	440	72.7	2522	20	AAW33729	Photorhabdus lumin
9	418	69.1	573	18	AAW18303	Photorhabdus lumin
10	418	69.1	573	19	AAW56559	Toxin TcdAIII, enc
11	418	69.1	2504	19	AAW56557	Toxin TcdA, encode
12	418	69.1	2504	22	AAW72610	Photorhabdus tcdA
13	418	69.1	2505	22	AAW72612	Modified Photorhab
14	410	67.8	2504	18	AAW17871	Photorhabdus lumin
15	395.5	65.4	2376	22	AAW97694	SEP4 protein encod
16	268	44.3	1565	18	AAW18305	Photorhabdus lumin
17	268	44.3	1565	19	AAW56568	Toxin TcdB, encode
18	174.5	28.8	562	18	AAW17886	Photorhabdus lumin
19	174.5	28.8	562	19	AAW56545	Toxin TcdAIII, enc
20	174.5	28.8	1189	18	AAW17884	Photorhabdus lumin
21	174.5	28.8	1189	19	AAW56543	Toxin TcdA, encode
22	79	13.1	250	22	ABB67792	Drosophila melanog
23	73.5	12.1	173	22	AAU25690	G protein-coupled
24	70.5	11.7	507	17	AAW00290	Mature Penicillin
25	70.5	11.7	532	17	AAW00291	Full length Penicl
26	69.5	11.5	467	21	AAW01601	Neisseria meningit
27	69.5	11.5	467	21	AAW75572	Neisseria meningit
28	69.5	11.5	467	22	AAW68934	Neisseria meningit
29	69.5	11.5	467	22	AAW68935	Neisseria meningit
30	69.5	11.5	601	23	ABF28734	Streptococcus poly
31	69.5	11.5	687	19	AAW41586	Truncated restin p
32	69.5	11.5	1392	20	AAW06999	Restin protein seq
33	69.5	11.5	1427	12	AAW10534	Human 160kD mediat
34	69	11.4	254	21	AAW29613	Cat flea HMT malvo
35	68	11.6	582	22	AAU03896	G protein-coupled
36	68	11.2	696	22	AAU33947	Staphylococcus aur
37	68	11.2	698	22	AAU36732	Staphylococcus aur
38	68	11.2	1056	23	ABP35622	Fungal ZBC protein
39	67.5	11.2	330	22	AAE01417	Human secreted pro
40	67.5	11.2	416	18	AAW55613	H. pylori ORF hp4e
41	67.5	11.2	635	18	AAW55723	H. pylori ORF hp4e
42	67.5	11.2	646	18	AAW55314	H. pylori ORF hp4e
43	67.5	11.2	668	18	AAW55709	H. pylori ORF hp4e
44	67.5	11.2	680	19	AAW46516	Saccharomyces cere
45	67.5	11.2	704	20	AAW17210	H. pylori outer me

ALIGNMENTS

RESULT 1
AAW18301 AAW18301 standard; Protein; 579 AA.
XX AC AAW18301;
XX DT 29-JAN-1998 (first entry)
XX DE Photorhabdus luminescens insect toxin TcdAIII.
XX KW Insecticide; insect; toxin; pest control; biological control;
KW Photorhabdus luminescens; TcdA; Southern corn rootworm;
KW Colorado potato beetle; Western corn rootworm; meal worm;
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
KW Diptera, Dictyoptera; Acarina; Homoptera.
XX OS Photorhabdus luminescens strain W-14 (ATCC 55397).

XX W09717432-A1.
 XX 15-MAY-1997.
 XX 06-NOV-1996; 96WO-US18003.
 XX 28-AUG-1996; 96US-0705484.
 XX 06-NOV-1995; 95US-0007255.
 XX 28-FEB-1996; 96US-0608423.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
 XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 XX Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 XX Strickland JA;
 XX WPI; 1997-281022/25.
 XX N-PSDB; AAT68845.
 XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
 XX can be genetically engineered into insect larvae food and plants for
 XX insect control
 XX Claim 34; Page 213-215; 276pp; English.
 XX This polypeptide comprises the 63.6 kDa TcdAIII insecticidal toxin
 XX protein of Photorhabdus luminescens W-14. Its sequence was deduced
 XX from a genomic DNA clone (AAT68845) and includes N-terminal and
 XX tryptic peptide sequences obtained from the isolated protein.
 XX TcdAIII is a proteolytic cleavage product of TcdA (see AAW17899).
 XX Claimed toxins of P. luminescens (see AAW17871, AAW17884-89,
 XX AAW17899-900, AAW18301-06) can be produced by recombinant DNA methods and
 XX applied to, or genetically engineered into, insect larvae food and
 XX plants for insect control. The toxins are particularly effective
 XX against Southern corn rootworm, Colorado potato beetle, Western corn
 XX rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet
 XX armyworm, black cutworm, cabbage looper, codling moth, corn earworm,
 XX European corn borer, tobacco hornworm and tobacco budworm
 XX (Lepidoptera), and are also active against insects of the orders
 XX Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All
 XX claimed).
 XX Sequence 579 AA;
 Alignment Scores:
 Pred. No.: Length: 579
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 18 Gaps: 0
 US-09-856-221-4 (1-332) x AAW18301 (1-579)
 QY 3 GAAGCGCAGTATTGCAAAACATCTGGAACCCACAGGCACAACTCAGGCACAG 62
 DB 281 GluAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnSerGln 300
 QY 63 CTGGCCCTCTCAAGCAAAATTCAGCAATACACGCTGTATTAAGTGGCTGGCGGA 122
 DB 301 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuLysThrLeuArgGlyArg 320
 QY 123 TTGGCGGCTATTATTATCAGTTTATGACVTGGCTGTTCCTCGTGGTTCATGGCTGAA 182
 DB 321 LeuAlaIaIeIyrPheGlnPheIyrAspLeuAlaValAlaArgCysLeuMetAlaGlu 340
 QY 183 CAAACTACAGCATGATGAATTAACCATTAAGCTGTACGCTTCATTAAAGCCGGTGCCTGG 242
 DB 341 GlnAlaIyrArgIrrpGluLeuAsnAspSerAlaArgPheIleIysProGlyAlaIrrp 360
 QY 243 CATGGCAGTATTGCTGGTGTGTGTACAGGTGAACCTGTGATGCTGAATTTGGCAGCATG 302

Db 361 GlnGlyThrTyrAlaGlyLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 380
 QY 303 GAAAAAATATTTCGCAAAAGATGAACGG 332
 DB 381 GluAspAlaHisLeuLysArgAspLysArg 390
 RESULT 2
 AAW56574
 ID AAW56574 standard; Protein; 579 AA.
 XX AC AAW56574;
 XX 07-AUG-1998 (first entry)
 XX Toxin TcdAIII, encoded by the tcdA gene from genomic region tcd.
 XX Photorhabdus luminescens W-14; nematode; symbiotic;
 XX Heterorhabdus; tca; tcb; tcc; tcd; insecticidal activity; toxin;
 XX Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 XX Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
 XX mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 XX cabbage looper; codling moth; corn earworm; European corn borer;
 XX Tobacco hornworm; budworm.
 XX Photorhabdus luminescens.
 XX OS W09808932-A1.
 XX 05-MAR-1998.
 XX 05-MAY-1997; 97WO-US07657.
 XX 06-NOV-1996; 96WO-US18003.
 XX 28-AUG-1996; 96US-0705484.
 XX 06-NOV-1996; 96US-0743699.
 XX (DOWC) DOWELANCO.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
 XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 XX Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 XX Strickland JA, Sukhapinda K;
 XX WPI; 1998-179427/16.
 XX N-PSDB; AAW29931.
 XX Isolated toxins from Photorhabdus luminescens strains - useful for
 XX control of insect pests
 XX Claim 34; Pages 245-247; 321pp; English.
 XX The present sequence represents a protein named TcdAIII of the bacterium
 XX Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
 XX nematodes of the Heterorhabdus genus. The bacterium has at least 4
 XX distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
 XX produced from these regions that are associated with insecticidal
 XX activity. The native toxins are secreted proteins. The proteins are
 XX toxic to insects upon exposure and especially when ingested. The
 XX nucleic acid sequence can be used to produce transgenic plants. The
 XX baculoviruses or microbial hosts for toxin production. They can be used
 XX to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 XX Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 XX Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 XX boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 XX codling moth, corn earworm, European corn borer or tobacco hornworm
 XX or budworm.
 XX Sequence 579 AA;
 Alignment Scores:
 Pred. No.: 1.77e-50 Length: 579

Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56574 (1-579)

QY 3 GAAGCGCAGTATTGCAAAAAAATCTGGAACCCACAGGCACAACTCAGGCACAG 62
 Db 281 GluAlaAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 300
 QY 63 CTGGCCCTTCTACAAAGCAATTCAGCAATACAGCTGTATTAAGTGGCTACGTGGCGCA 122
 Db 301 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyraAsnTrpLeuArgGlyArg 320
 QY 123 TTGGCGCGCTATTATATACGTTTATGACTGGCTGGTGTCCCTGTTGTTGATGGCTGAA 182
 Db 321 LeuAlaAlaIleTyPheGlnPheTyAspLeuAlaValAlaArgCysLeuMetAlaGlu 340
 QY 183 CAAACTTACCAGTATGAATGAAGCAATAAGCTGTACGCTTCATTAAGCCCGGTGCTGG 242
 Db 341 GlnAlaTyArgTrpGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 360
 QY 243 CATGGCAGTATGCTGCTTGTAGCAGTGAACTTCATGCTGCAATTTGGCACAGATG 302
 Db 361 GlnGlyThrTyraGlyLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 380
 QY 303 GAAAAAACTATTGGAAGAAAGATGAACGG 332
 Db 381 GluAspAlaHisLeuLysArgAspLysArg 390

RESULT 3

ID AAW17899 standard; Protein: 2516 AA.

AC AAW17899;

DT 29-JAN-1998 (first entry)

DE Photorhabdus luminescens insect toxin TcdA.

XX Insecticide; insect; toxin; pest control; biological control;
 KW Photorhabdus luminescens; TcdA; Southern corn rootworm;
 KW Colorado potato beetle; Western corn rootworm; meal worm;
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
 KW Diptera, Dictyoptera; Acarina; Homoptera.

OS Photorhabdus luminescens strain W-14 (ATCC 55397).

Key Location/Qualifiers
 FT Protein 1..2516
 FT /label= TcdA
 FT Protein 89..1937
 FT /label= TcdAii
 FT Peptide 89..100
 FT /note= "S2 N-terminus (Claim 30)"
 FT Peptide 284..299
 FT /note= "tryptic peptide (Claim 30)"
 FT Peptide 554..563
 FT /note= "tryptic peptide (Claim 30)"
 FT Peptide 1080..1092
 FT /note= "isolated N-terminal peptide (Claim 30)"
 FT Peptide 1385..1400
 FT /note= "tryptic peptide (Claim 30)"
 FT Peptide 1478..1497
 FT /note= "tryptic peptide (Claim 30)"
 FT Peptide 1620..1642
 FT /note= "claimed peptide (Claim 30)"
 FT Peptide 1938..1948
 FT /note= "N-terminal peptide (Claim 30)"

FT Protein 1938..2516
 FT /label= TcdAii
 FT Peptide 2327..2345
 FT /note= "tryptic peptide (Claim 30)"
 FT Peptide 2398..2408
 FT /note= "tryptic peptide (Claim 30)"
 PN WO9717432-A1.
 XX 15-MAY-1997.
 XX 06-NOV-1996; 96WO-US18003.
 XX 28-AUG-1996; 96US-0705484.
 PR 06-NOV-1995; 95US-0007255.
 PR 28-FEB-1996; 96US-0608423.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
 PI Firench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA;
 XX WPI; 1997-281022/25.
 DR N-PSDB; AAT68843.
 XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
 PT can be genetically engineered into insect larvae food and plants for
 PT insect control
 PT Claim 34; page 185-194; 276pp; English.
 XX This polypeptide comprises the 282.9 kDa TcdA insecticidal toxin
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
 CC from a genomic DNA clone (AAT68843) and includes N-terminal and
 CC tryptic peptide sequences obtained from the isolated protein. The
 CC 282.9 kDa insect toxin is proteolytically processed into TcdAii and
 CC TcdAiii components (see AAW17900 and AAW18301). Claimed toxins of P.
 CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can
 CC be produced by recombinant DNA methods and applied to, or genetically
 CC engineered into, insect larvae food and plants for insect control.
 CC The toxins are particularly effective against Southern corn rootworm,
 CC Colorado potato beetle, Western corn rootworm, meal worm, boll weevil
 CC and turf grub (Coleoptera), beet armyworm, black cutworm, cabbage
 CC looper, codling moth, corn earworm, European corn borer, tobacco
 CC hornworm and tobacco budworm (Lepidoptera), and are also active
 CC against insects of the orders Hymenoptera, Diptera, Dictyoptera,
 CC Acarina and Homoptera. (All claimed).

SQ Sequence 2516 AA;

Alignment Scores:
 Pred. No.: 2.75e-50 Length: 2516
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 18 Gaps: 0
 US-09-856-221-4 (1-332) x AAW17899 (1-2516)
 QY 3 GAAGCGCAGTATTGCAAAAAAATCTGGAACCCACAGGCACAACTCAGGCACAG 62
 Db 2218 GluAlaAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2237
 QY 63 CTGGCCCTTCTACAAAGCAATTCAGCAATACAGCTGTATTAAGTGGCTACGTGGCGCA 122
 Db 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyraAsnTrpLeuArgGlyArg 2257
 QY 123 TTGGCGCGCTATTATATACGTTTATGACTGGCTGTTTCCCTGTTGTTGATGGCTGAA 182
 Db 2258 LeuAlaAlaIleTyPheGlnPheTyAspLeuAlaValAlaArgCysLeuMetAlaGlu 2277

QY 183 CAACCTTACCAGTATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGCTGG 242
 DB 2278 GlnAlaTyrArgTrpGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 2297
 QY 243 CATGGCACTTATCTGCTGGTTTGTACGAGTGAACCTTATGCTGAATTTGGCGACAGT 302
 DB 2298 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317
 QY 303 GAAAAAACAATTTGAAAAAGATGAACGG 332
 DB 2318 GluAspAlaHisLeuLysArgAspLysArg 2327
 RESULT 4
 AAW56572
 ID AAW56572 standard; Protein: 2516 AA.
 AC AAW56572;
 DT 07-AUG-1998 (first entry)
 DE Toxin TcdA, encoded by the tcdA gene from genomic region tcd.
 KW Photobacterium luminescens W-14; nematode; symbiotic;
 KW Heterorhabdus; tca; tcb; tcc; tcd; insecticidal activity; toxin;
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW Tobacco hornworm; budworm.
 OS Photobacterium luminescens.
 PN W09808932-A1.
 PD 05-MAR-1998.
 PF 05-MAY-1997; 97WO-US07657.
 PR 06-NOV-1996; 96WO-US18003.
 PR 28-AUG-1996; 96US-0705484.
 PR 06-NOV-1996; 96US-0743699.
 PA (DOWC) DOWELANCO.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Patig R;
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA, Sukhapinda K;
 WPI; 1998-179427/16.
 N-PSDB; AAV29928.
 Isolated toxins from Photobacterium luminescens strains - useful for control of insect pests
 Claim 34; Pages 224-231; 321pp; English.
 The present sequence represents a protein named TcdA of the bacterium Photobacterium luminescens (W-14). This is a symbiotic bacterium of the nematodes of the Heterorhabdus genus. The bacterium has at least 4 distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are produced from these regions that are associated with insecticidal activity. The native toxins are secreted proteins. The proteins are toxic to insects upon exposure and especially when ingested. The nucleic acid sequence can be used to produce transgenic plants, baculoviruses or microbial hosts for toxin production. They can be used to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera, Dictyoptera, Acarina or Homoptera orders, especially the Southern or Western corn rootworm, Colorado potato beetle, mealworm, boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth, corn earworm, European corn borer or tobacco hornworm

CC or budworm.
 SQ Sequence 2516 AA;
 Alignment Scores:
 Pred. No.: 2,75e-50 Length: 2516
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 19 Gaps: 0
 US-09-856-221-4 (1-332) x AAW56572 (1-2516)
 QY 3 GAAGCGCAGTATTCGAAAAAACTATCTGAAACCCAGCAGCACAACCTCAGGCACAG 62
 DB 2218 GluAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2237
 QY 63 CTGGCCTTCCTCAAGCAAAATTCAGCAATACAGCGTGTATTAACCTGGCTAGCGCGCA 122
 DB 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2257
 QY 123 TTGGCGCTATTATTATTCAGTCTTTATGACTTGCTGTTTCCCTGTGTTGATGGCTGAA 182
 DB 2258 LeuAlaAlaIleTyrPheGlnPheTyrAspLeuAlaValAlaArgCysLeuMetAlaGlu 2277
 QY 183 CAACCTTACCAGTATGAATTTGAACGATGAACGCTGTACGCTTCATTAAAGCCGCTGCTGG 242
 DB 2278 GlnAlaTyrArgTrpGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 2297
 QY 243 CATGGCACTTATCTGCTGGTTTGTACGAGTGAACCTTATGCTGAATTTGGCGACAGT 302
 DB 2298 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317
 QY 303 GAAAAAACAATTTGAAAAAGATGAACGG 332
 DB 2318 GluAspAlaHisLeuLysArgAspLysArg 2327
 RESULT 5
 AAB72609
 ID AAB72609 standard; Protein: 2516 AA.
 AC AAB72609;
 DT 04-MAY-2001 (first entry)
 DE Photobacterium tcdA toxin.
 KW TcdA; TcdA; insect toxin; plant; insect resistance.
 OS Photobacterium sp.
 PN W0200111029-A1.
 PD 15-FEB-2001.
 PF 11-AUG-2000; 2000WO-US22237.
 PR 11-AUG-1999; 99US-0148356.
 PA (DOWC) DOW AGROSCIENCES LLC.
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 WPI; 2001-191536/19.
 N-PSDB; AAF58778.
 Novel polynucleotide sequence encoding insect toxins, useful for producing transgenic plants having resistance to insects, especially corn rootworm
 Disclosure; Page 51-61; 106pp; English.

Not prior art

XX The present invention provides the protein and coding sequences of
 CC modified versions of the Photorhabdus TcdA and TcdB toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC Photorhabdus TcdA protein.

XX SQ Sequence 2516 AA;

Alignment Scores:
 Pred. No.: 2,75e-50 Length: 2516
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 22 Gaps: 0

US-09-856-221-4 (1-332) x AAB72609 (1-2516)

QY 3 GAAGCGCAGTATTGCAAAAACCTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
 Db 2218 GluAlaAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2237
 QY 63 CTGGCCCTTCTCAAGAGCAATTCAGCAATACAGCGTTCTATACTAGCTGGCTGGCGCA 122
 Db 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2257
 QY 123 TTGGCGGCTATTATATACGTTTATGACTTGCGTGTTCCTGTTGCTGTTGCTGAA 182
 Db 2258 LeuAlaAlaIleTyrPheGlnPheTyrAspLeuAlaValAlaArgCysLeuMetAlaGlu 2277
 QY 183 CAAACTTACCAGTATGAATTTGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGCTG 242
 Db 2278 GlnAlaTyrArgTrpGlnLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 2297
 QY 243 CATGGCACTATGCTGGTTGTGTAGCAGGTGAACCTTGCTGCTGAATTTGGCAGCAGATG 302
 Db 2298 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317
 QY 303 GAAAAAACTATTGGAAGAGTGAACGG 332
 Db 2318 GluAspAlaHisLeuLysArgAspLysArg 2327

RESULT 6

AAAB72611
 ID AAB72611 standard; Protein: 2517 AA.

AC AAB72611;

DT 04-MAY-2001 (first entry)

DE Modified Photorhabdus tcdA toxin.

KW TcdA; TcdB; insect toxin; plant; insect resistance.

OS Photorhabdus sp.

XX Synthetic.

PN WO200111029-A1.

PD 15-FEB-2001.

PF 11-AUG-2000; 2000WO-US22237.

PR 11-AUG-1999; 99US-0148356.

XX (DOWC) DOW AGROSCIENCES LLC.

PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

PI Sukhapinda K, Merlo AO;

XX WPI; 2001-191536/19.

DR N-PSDB; AAF58780.

XX Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 XX corn rootworm.

PS Claim 1; Page 72-83; 106pp; English.

XX The present invention provides the protein and coding sequences of
 CC modified versions of the Photorhabdus TcdA and TcdB toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC modified Photorhabdus TcdA protein.

XX SQ Sequence 2517 AA;

Alignment Scores:
 Pred. No.: 2,75e-50 Length: 2517
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 22 Gaps: 0

US-09-856-221-4 (1-332) x AAB72611 (1-2517)

QY 3 GAAGCGCAGTATTGCAAAAACCTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
 Db 2219 GluAlaAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2238
 QY 63 CTGGCCCTTCTCAAGAGCAATTCAGCAATACAGCGTTCTATACTAGCTGGCTGGCGCA 122
 Db 2239 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2258
 QY 123 TTGGCGGCTATTATATACGTTTATGACTTGCGTGTTCCTGTTGCTGTTGCTGAA 182
 Db 2259 LeuAlaAlaIleTyrPheGlnPheTyrAspLeuAlaValAlaArgCysLeuMetAlaGlu 2278
 QY 183 CAAACTTACCAGTATGAATTTGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGCTG 242
 Db 2279 GlnAlaTyrArgTrpGlnLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 2298
 QY 243 CATGGCACTTATGCTGGTTGTGTAGCAGGTGAACCTTGCTGCTGAATTTGGCAGCAGATG 302
 Db 2299 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2318
 QY 303 GAAAAAACTATTGGAAGAGTGAACGG 332
 Db 2319 GluAspAlaHisLeuLysArgAspLysArg 2328

RESULT 7

AAAB72614

ID AAB72614 standard; Protein: 2537 AA.

AC AAB72614;

DT 04-MAY-2001 (first entry)

XX TcdA toxin-zein ER signal peptide fusion protein.

DE TcdA; TcdB; insect toxin; plant; insect resistance.

KW TcdA; TcdB; insect toxin; plant; insect resistance.

OS Chimeric - Photorhabdus sp.

OS Chimeric - Zea mays.

PN WO200111029-A1.

PD 15-FEB-2001.

PF 11-AUG-2000; 2000WO-US22237.

XX 11-AUG-1999; 99US-0148356.

XX (DOWC) DOW AGROSCIENCES LLC.
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 XX WPI: 2001-191536/19.
 DR N-PSDB: AAF58783.
 XX Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm
 XX Example 1: Page 94-104; 106pp; English.
 XX The present invention provides the protein and coding sequences of
 CC modified versions of the Photobabidus TcdA and TcdB toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is a
 CC fusion protein comprising the Photobabidus TcdA toxin and the maize zein
 CC protein ER signal peptide.
 XX SQ Sequence 2537 AA;
 Alignment Scores:
 Pred. No.: 2,76e-50 Length: 2537
 Score: 446.00 Matches: 86
 Percent Similarity: 89.09% Conservative: 12
 Best Local Similarity: 78.18% Mismatches: 12
 Query Match: 73.72% Indels: 0
 DB: 22 Gaps: 0
 US-09-856-221-4 (1-332) x AAB72614 (1-2537)
 QY 3 GAAGCGCAGTATTGCAAAAATCTCTGCAACCCACAGCACAACCTCAGGCACAG 62
 Db 2239 GluAlaAlaValleuGlnlyThrSerLeuLysThrGlnGlnGlnGlnGlnGln 2258
 QY 63 CTGGCCTCTCTCAACAGCAATACAGCGTGTGATACCTGGCTACGTTGGCGGA 122
 Db 2259 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2278
 QY 123 TTGGCGGCTATTATTATCAGTTTATGACTTGCTGCTTTCCCTGTTGATGCTGAA 182
 Db 2279 LeuAlaAlaLeuPheGlnPheThrAspLeuAlaValAlaArgCysLeuMetAlaGlu 2298
 QY 183 CAACCTACAGCATGAATGAACGATTAACCTGCTAGCTTCATTAAAGCCGGTGCCTGG 242
 Db 2299 GlnAlaTyrArgTrpGluLeuAsnAspSerAlaArgPheLeuLysProGlyAlaTrp 2318
 QY 243 CATGGCCTATGCTGCTGTTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGATG 302
 Db 2319 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2338
 QY 303 GAATAAATCTATTGGAAGAAGATGAACGG 332
 Db 2339 GluAspAlaHisLeuLysArgAspLysArg 2348
 RESULT 8
 AAY33729
 ID AAY33729 standard; Protein: 2522 AA.
 XX
 XX AAY33729;
 XX
 XX 09-NOV-1999 (first entry)
 DT Photobabidus luminescens hph2-encoded insecticidal toxin.
 DE
 XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
 KW
 XX Photobabidus luminescens.
 OS
 XX

PN W09942589-A2.
 XX
 XX 26-AUG-1999.
 PD
 XX
 PF 18-FEB-1999; 99WO-EP01015.
 XX
 XX 20-JAN-1999; 99US-0116439.
 PR 20-FEB-1998; 98US-0027080.
 XX
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX
 XX Anderson AR, Chen JS, Dunn MM, Hart HP, Kramer VC;
 PI Morgan MK, Warren GW;
 PI
 DR WPI: 1999-527479/44.
 DR N-PSDB: AAZ06831.
 XX
 XX New nucleic acid from Photobabidus luminescens encoding insecticidal
 PT toxins, used for making resistant transgenic plants
 PT
 XX Claim 26; Page 133-140; 148pp; English.
 XX
 XX This sequence represents the hph2 gene encoded insecticidal
 CC toxin from Photobabidus luminescens. It is one of three insecticidal
 CC toxins (AAY33728-Y33730) encoded by open reading frames (orfs) in a 38kb
 CC fragment of P. luminescens DNA (AAZ06831). The hph2 gene was identified
 CC using probe #2 which was amplified from P. luminescens genomic DNA using
 CC PCR primers AAZ06829-206830. P. luminescens is a member of the
 CC Enterobacteriaceae family and is a symbiotic bacterium of
 CC nematodes of the genus Heterorhabditis. The nematodes colonise
 CC insect larvae, kill them, and their offspring feed on the dead
 CC larvae. However, the insecticidal agents are produced by P.
 CC luminescens rather than the nematodes. The toxins have activity against
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni),
 CC European Corn Borer (Ostrinia nubilalis) and Fall Armyworm
 CC (Spodoptera frugiperda) and also against Coleopteran insects
 CC (e.g., Colorado potato beetle, Leptinotarsa decemlineata). In
 CC addition the toxins are active against strains resistant to known
 CC insecticides. The DNA sequence can be used to generate transgenic plants
 CC of various species that are resistant to economically important insect
 CC pests and also for recombinant production of the toxins for use as
 CC insecticides.
 XX SQ Sequence 2522 AA;
 Alignment Scores:
 Pred. No.: 1,75e-49 Length: 2522
 Score: 440.00 Matches: 84
 Percent Similarity: 88.18% Conservative: 13
 Best Local Similarity: 76.36% Mismatches: 13
 Query Match: 72.73% Indels: 0
 DB: 20 Gaps: 0
 US-09-856-221-4 (1-332) x AAY33729 (1-2522)
 QY 3 GAAGCGCAGTATTGCAAAAATCTCTGCAACCCACAGCACAACCTCAGGCACAG 62
 Db 2224 GluAlaAlaValleuGlnlyThrSerLeuLysThrGlnGlnGlnGlnGlnGln 2243
 QY 63 CTGGCCTCTCTCAACAGCAATACAGCGTGTGATACCTGGCTACGTTGGCGGA 122
 Db 2244 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2263
 QY 123 TTGGCGGCTATTATTATCAGTTTATGACTTGCTGCTTTCCCTGTTGATGCTGAA 182
 Db 2264 LeuAlaAlaLeuPheGlnPheThrAspLeuAlaLeuAlaArgCysLeuMetAlaGlu 2283
 QY 183 CAACCTACAGCATGAATGAACGATTAACCTGCTAGCTTCATTAAAGCCGGTGCCTGG 242
 Db 2284 GlnAlaTyrArgTrpGluLeuAspSerAlaArgPheLeuLysProGlyAlaTrp 2303
 QY 243 CATGGCCTATGCTGCTGTTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGATG 302

Not for use

|||||
 Db 2304 GlnGlyThrTyraGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2323
 QY 303 GAAAAAACTATTGGGAAAGATGAACGG 332
 Db 2324 GluAspAlaHisLeuArgAspLysArg 2333
 RESULT 9
 AAW18303
 ID AAW18303 standard; Protein: 573 AA.
 XX
 AC AAW18303;
 XX
 DT 30-JAN-1998 (first entry)
 XX
 DE Photorhabdus luminescens insect toxin TcbAIII.
 XX
 KW Insecticide; insect; toxin; pest control; biological control;
 KW Photorhabdus luminescens; TcbA; Southern corn rootworm;
 KW Colorado potato beetle; Western corn rootworm; meal worm;
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
 KW Diptera, Dictyoptera; Acarina; Homoptera.
 XX
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).
 XX
 PN W09717432-A1.
 XX
 PD 15-MAY-1997.
 XX
 PF 06-NOV-1996; 96WO-US18003.
 XX
 PR 28-AUG-1996; 96US-0705484.
 PR 06-NOV-1995; 95US-0007255.
 PR 28-FEB-1996; 96US-0608423.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA;
 XX
 DR WPI: 1997-281022/25.
 DR N-PSDB: AAT68847.
 XX
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
 PT can be genetically engineered into insect larvae food and plants for
 PT insect control
 XX
 PS Claim 34; Page 229-231; 276pp; English.
 CC
 CC This polypeptide comprises the 62.9 kDa TcbAIII insecticidal toxin
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced
 CC from a genomic DNA clone (AAV68847). TcbAIII is a proteolytic
 CC cleavage product of TcbA (see AAW17871). Claimed toxins of P.
 CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can
 CC be produced by recombinant DNA methods and applied to, or genetically
 CC engineered into, insect larvae food and plants for insect control.
 CC The toxins are particularly effective against Southern corn rootworm,
 CC Colorado potato beetle, Western corn rootworm, meal worm, boll
 CC weevil and turf grub (Coleoptera), beet armyworm, black cutworm,
 CC cabbage looper, codling moth, corn earworm, European corn borer,
 CC tobacco hornworm and tobacco budworm (Lepidoptera), and are also
 CC active against insects of the orders Hymenoptera, Diptera,
 CC Dictyoptera, Acarina and Homoptera. (All claimed).
 XX
 SQ Sequence 573 AA;
 Alignment Scores:
 Pred. No.: 9.83e-47 Length: 573
 Score: 418.00 Matches: 78

Percent Similarity: 84.55% Conservative: 15
 Best Local Similarity: 70.91% Mismatches: 17
 Query Match: 69.09% Indels: 0
 DB: 18 Gaps: 0
 US-09-856-221-4 (1-332) x AAW18303 (1-573)
 QY 3 GAAGCGCAGTATTGCAAAAAAACTATCTGGAACCAACAGGACAACTCAGGCACAG 62
 |||||
 Db 281 GluAlaAlaGluMetGlnLysLysLysLeuLysThrGlnGlnAlaGlnAlaGln 300
 |||||
 QY 63 CTGGCTTCCTACAAAGCAAAATTCAGCAATACAGCTGTATTAACCTGCTACCTGGCCCA 122
 |||||
 Db 301 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTySerTrpLeuArgGlyArg 320
 |||||
 QY 123 TTGGCGGCTATTATATACATTTTATGACTTGGCTGCTGCTGCTGCTGCTGCTGCTGAA 182
 |||||
 Db 321 LeuSerGlyIleTyPheGlnPheTyAspLeuAlaValSerA-gCysLeuMetAlaGlu 340
 |||||
 QY 183 CAAACTTACACATATGAATTCAGCATAAAGCTGTACGCTTCATTAAAGCCGCTGCTGG 242
 |||||
 Db 341 GlnSerTyGlnTrpGluAlaAsnAspAsnSerIleSerPheValLysProGlyAlaTrp 360
 |||||
 QY 243 CATGGCACTTATGCTGTTTGTAGCAGGTGAACCTTGCCTCAATTTGGCACAGATG 302
 |||||
 Db 361 GlnGlyThrTyAlaGlyLeuLeuGlyGluAlaLeuIleGlnAsnLeuAlaGlnMet 380
 |||||
 QY 303 GAAAAAACTATTGGAAGATGAACGG 332
 |||||
 Db 381 GluGluAlaTyLeuLysTrpGluSerArg 390
 RESULT 10
 AAW56559
 ID AAW56559 standard; Protein: 573 AA.
 XX
 AC AAW56559;
 XX
 DT 07-AUG-1998 (first entry)
 XX
 DE Toxin TcbAIII, encoded by the tcbA gene from genomic region tcb.
 KW Photorhabdus luminescens W-14; nematode; symbiotic;
 KW Heterorhabditis; tca; tcb; tcc; insecticidal activity; toxin;
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 KW Homoptera; Southern Western corn rootworm; Colorado potato beetle;
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW Tobacco hornworm; budworm.
 XX
 OS Photorhabdus luminescens.
 XX
 PN W09808932-A1.
 PD
 PD 05-MAR-1998.
 XX
 PF 05-MAY-1997; 97WO-US07657.
 XX
 PR 06-NOV-1996; 96WO-US18003.
 PR 28-AUG-1996; 96US-0705484.
 PR 06-NOV-1996; 96US-0743699.
 XX
 PA (DOMC) DOWELANCO.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA, Sukhinda K;
 XX
 DR WPI: 1998-179427/16.
 DR N-PSDB: AAV29987.
 XX
 PT Isolated toxins from Photorhabdus luminescens strains - useful for

PT control of insect pests

PS Claim 34; Pages 261-263; 321pp; English.

XX The present sequence represents a protein named TcbA11 of the bacterium
 CC Photobacterium luminescens (W-14). This is a symbiotic bacterium of the
 CC nematodes of the Heterorhabdus genus. The bacterium has at least 4
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
 CC produced from these regions that are associated with insecticidal
 CC activity. The native toxins are secreted proteins. The proteins are
 CC toxic to insects upon exposure and especially when ingested. The
 CC nucleic acid sequence can be used to produce transgenic plants,
 CC baculoviruses or microbial hosts for toxin production. They can be used
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 CC codling moth, corn earworm, European corn borer or tobacco hornworm
 CC or budworm.

XX Sequence 573 AA;

Alignment Scores:

Pred. No.: 9.82e-47 Length: 573
 Score: 418.00 Matches: 78
 Percent Similarity: 84.55% Conservative: 15
 Best Local Similarity: 70.91% Mismatches: 17
 Query Match: 69.09% Indels: 0
 DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56559 (1-573)

QY 3 GAAGCGCGAGTATTGCAAAAAAATCTCTGGAACCCACAGGCACAACTCAGGCACAG 62
 Db 281 GLUAlaAlaGluMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 300
 QY 63 CTGGCTTCTCTAAAGCAAAATTCAGCAATACAGCGTTGTATACGTGCTAGTGGCGCA 122
 Db 301 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 320
 QY 123 TTGGCGGCTATTAT 182
 Db 321 LeuSerGlyLeuTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 340
 QY 183 CAACACTTACCAGTATGATTAAGCAATAAGCTGTACGCTTCATTAAAGCCCGGTGCTGG 242
 Db 341 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerIleSerPheValLysProGlyAlaTrp 360
 QY 243 CATGGCACTATGCTGGTTGTGTAGCAGGTCAACCTTGATGCTGATTTGGCAGCATG 302
 Db 361 GlnGlyThrTyrAlaGlyLeuLeuCysGlyGluAlaLeuIleGlnAsnLeuAlaGlnMet 380
 QY 303 CAATAAACTATTGTAAGAAAGATGAACGG 332
 Db 381 GluGluAlaTyrLeuLysTrpGluSerArg 390

RESULT 11

AAW56557

ID AAW56557 standard; Protein; 2504 AA.

XX AAW56557;

07-AUG-1998 (first entry)

Toxin TcbA, encoded by the tcbA gene from genomic region tcb.

XX Photobacterium luminescens W-14; nematode; symbiotic;
 KW Heterorhabdus; tca; tcb; tcc; tcd; insecticidal activity; toxin;
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
 KW cabbage looper; codling moth; corn earworm; European corn borer;
 KW Tobacco hornworm; budworm.

XX Photobacterium luminescens.
 OS W09808932-A1.
 PN 05-MAR-1998.
 PD 05-MAY-1997; 97WO-US07657.
 PF 06-NOV-1996; 96WO-US18003.
 PR 28-AUG-1996; 96US-0705484.
 PR 06-NOV-1996; 96US-0743699.
 PA (DOWC) DOWELANCO.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
 PI Petrelli J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA, Sukhapinda K;
 XX WPI; 1998-179427/16.
 DR N-PSDB; AAV29985.
 XX Isolated toxins from Photobacterium luminescens strains - useful for
 PT control of insect pests
 XX Claim 34; Pages 163-169; 321pp; English.
 CC The present sequence represents a protein named TcbA of the bacterium
 CC Photobacterium luminescens (W-14). This is a symbiotic bacterium of the
 CC nematodes of the Heterorhabdus genus. The bacterium has at least 4
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
 CC produced from these regions that are associated with insecticidal
 CC activity. The native toxins are secreted proteins. The proteins are
 CC toxic to insects upon exposure and especially when ingested. The
 CC nucleic acid sequence can be used to produce transgenic plants,
 CC baculoviruses or microbial hosts for toxin production. They can be used
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 CC codling moth, corn earworm, European corn borer or tobacco hornworm
 CC or budworm.

XX Sequence 2504 AA;

Alignment Scores:

Pred. No.: 1.53e-46 Length: 2504
 Score: 418.00 Matches: 78
 Percent Similarity: 84.55% Conservative: 15
 Best Local Similarity: 70.91% Mismatches: 17
 Query Match: 69.09% Indels: 0
 DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56557 (1-2504)

QY 3 GAAGCGCGAGTATTGCAAAAAAATCTCTGGAACCCACAGGCACAACTCAGGCACAG 62

Db 2212 GLUAlaAlaGluMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 2231

QY 63 CTGGCTTCTCTAAAGCAAAATTCAGCAATACAGCGTTGTATACGTGCTAGTGGCGCA 122

Db 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 2251

QY 123 TTGGCGGCTATTAT 182

Db 2252 LeuSerGlyLeuTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 2271

QY 183 CAACACTTACCAGTATGATTAAGCAATAAGCTGTACGCTTCATTAAAGCCCGGTGCTGG 242

Db 2272 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerIleSerPheValLysProGlyAlaTrp 2291

DR WPI: 2001-169009/17.
 DR N-PSDB: AAA91292.
 XX
 PT New nucleic acid encoding a polypeptide useful as a pesticide
 PT especially for Coleoptera -
 XX
 XX Claim 24; Page 92-99; 109pp; English.
 PS
 CC This sequence represents the SepA protein encoded by the Serratia
 CC insecticidal protein complex gene of the invention. The invention relates
 CC to a gene encoding an insecticidal protein complex or a functional
 CC fragment, a neutral mutation, or a homologue of the complex. The
 CC polypeptides and nucleotides of the invention are used to induce amber
 CC diseases or like conditions in insects, especially to those from the
 CC order comprising Coleoptera, useful as a pesticide.
 XX
 SQ Sequence 2376 AA:

Alignment Scores:
 Pred. No.: 1.54e-43 Length: 2376
 Score: 395.50 Matches: 80
 Percent Similarity: 82.73% Conservative: 11
 Best Local Similarity: 72.73% Mismatches: 18
 Query Match: 65.37% Indels: 1
 DB: 22 Gaps: 1

US-09-856-221-4 (1-332) x AAY97694 (1-2376)

QY 3 GAAGCGCAGTATTGCAAAAACACTATCTGGAACCCACAGCACAACTCAGGCACAG 62
 DB 2087 GluGlyAlaGluLeuGlnTyrThrLeuGluThrGlnGlnAlaGlnAlaGln 2106
 QY 63 CTGGCCTTCCTACAAAGCAATTCAGCAATACAGCGTCTATACCTGCTACGTGGCGA 122
 DB 2107 LeuAlaPheLeuGlnSerLysPheAsnThrAlaLeuTyrSerIrrPleuArgGlyArg 2126
 QY 123 TTGGCGGCTATTATTATACAGTTTATGACTTTGGCTGTTCCCTGTGTGTGATGGCTGAA 182
 DB 2127 LeuSerAlaIleTyrTyrGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGln 2146
 QY 183 CAAACTTACCAGTATGAATTGAACGATGAAGCTGACGCTTCATTAAGCCGCGTGGCTGG 242
 DB 2147 GlnAlaTrpGlnTrpAsp---LysPheGluThrArgSerPheIleGlnProGlyAlaTrp 2165
 QY 243 CATGGCATTATCTGCTTGTGTACAGGTGAACCTTGATGCTGAATTGGCACAGATG 302
 DB 2166 MetGlyAlaAsnAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuAsnLeuAlaGlnMet 2185
 QY 303 GAAAAAACTATTGCAAAAAGATGAACGG 332
 DB 2186 GluGlnAlaTrpLeuThrGlyAspGluArg 2195

Search completed: January 15, 2003, 15:50:46
 Job time : 39.7357 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 15, 2003, 15:39:33 : Search time 41.5673 Seconds
(without alignments)
2878.683 Million cell updates/sec

Title: us-09-856-221-1
Perfect score: 793
Sequence: 1 tctgcagcactcacttat.....gtgacccaactgatacagtt 449

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameter--

Minimum DB seq length: 0
Maximum DB seq length: 20000

Post-processing: Minimum Match
Maximum Match
Listing first

Command line parameters:
-MODEL=frame_n2p_model -DEV=xl
-O=/cgn2.1/USPTO.spool/US098562:
-DB=A_Geneseq_101002_OFNP-fasta
-LOOPEXT=0 UNITS=bits -START=1
-LIST=45 -DOALIGN=200 -THR_SCORE=1
-MODE=LOCAL -OUTFMT=ptc -NORM=ext
-USER=US09856221.ccn_1_1_60.0run
-NO_XLIFY -NO_MMAP -LARGEQUERY -NI
-WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPOP=10 -XGAPEXT=0.5 -XGAPOP=6 -XGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -XGAPOP=6 -XGAPEXT=7

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	434.5	54.8	2516	18	AAW17899	Photobabidus lumin
2	434.5	54.8	2516	19	AAW56572	Toxin TcdA, encode
3	434.5	54.8	2516	22	AAW72609	Toxin TcdA, encode
4	434.5	54.8	2517	22	AAW72611	Modified Photobab
5	434.5	54.8	2517	22	AAW72614	TcdA toxin-zeln ER
6	428.5	54.0	2522	20	AAW33729	Photobabidus lumin
7	416.5	52.5	2376	22	AAW97694	SepA protein encod
8	415	52.3	2504	19	AAW56557	Toxin TcdA, encode
9	415	52.3	2504	22	AAW72610	Photobabidus tcdA
10	412	52.0	2505	22	AAW72612	Modified Photobab
11	410	51.7	2504	18	AAW17871	Photobabidus lumin
12	354	44.6	579	18	AAW18301	Photobabidus lumin
13	354	44.6	579	19	AAW56574	Toxin TcdAIII, enc
14	353	44.5	573	18	AAW18303	Photobabidus lumin
15	353	44.5	573	19	AAW56559	Toxin TcdAIII, enc
16	227	28.6	1189	18	AAW17884	Photobabidus lumin
17	227	28.6	1189	19	AAW56543	Toxin TcdB, encode
18	208	26.2	1565	19	AAW56568	Photobabidus lumin
19	201	25.3	582	18	AAW17886	Toxin TcdB, encode
20	201	25.3	582	19	AAW56545	Photobabidus lumin
21	190.5	24.0	1565	18	AAW18305	Photobabidus lumin
22	93	11.7	1849	18	AAW17900	Photobabidus lumin
23	93	11.7	1849	19	AAW56573	Toxin TcdAIII, enco
24	84.5	10.7	568	22	AAW72972	Chlamydia pneumoni
25	84.5	10.7	568	23	ABB90550	Chlamydia pneumoni
26	81	10.2	578	22	ABB71138	Drosophila melanog
27	80.5	10.2	446	22	AAU63108	Propionibacterium
28	78.5	9.9	662	22	AAE00477	scal protein invol
29	77	9.7	3201	22	ABB62899	Drosophila melanog
30	76	9.6	841	22	ABG19955	Novel human diagno
31	76	9.6	1014	23	AAU96745	PC2-PPAR gamma co
32	76	9.6	1447	22	AAW86975	D. melanogaster la
33	76	9.6	2799	19	AAW81867	Human tumour suppr
34	74	9.3	1434	22	AAW83976	Amino acid sequenc
35	73.5	9.3	338	23	AAU77921	Human Teta-binding
36	73.5	9.3	371	20	AAW73369	Epitope tagged TBP
37	73.5	9.3	716	22	ABG28268	Novel human diagno
38	73	9.2	149	22	ABB64213	Drosophila melanog
39	72.5	9.1	805	22	ABB62304	Drosophila melanog
40	72.5	9.1	926	22	ABB62874	Drosophila melanog
41	72.5	9.1	4498	22	ABB58595	Drosophila melanog
42	72	9.1	1905	22	ABB59243	Drosophila melanog
43	72	9.1	1905	22	AAU38925	Drosophila G-proce
44	71.5	9.0	399	21	AAW34742	Human secreted pro
45	71.5	9.0	694	20	AAW17829	Human PKO394 prote

ALIGNMENTS

RESULT 1
AAW17899
ID AAW17899 standard; Protein; 2516 AA.

AC AAW17899;

DT 29-JAN-1998 (first entry)

DE Photobabidus luminescens insect toxin TcdA.

XX Insecticide; insect; pest control; biological control;
XX Photobabidus luminescens; TcdA; Southern corn rootworm;
XX Colorado potato beetle; Western corn rootworm; meal worm;
XX boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
XX cabbage looper; codling moth; corn earworm; European corn borer;
XX tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
XX Diptera, Dictyoptera; Acarina; Homoptera.
OS Photobabidus luminescens strain W-14 (ATCC 55397).

PA (DOWC) DOWELANCO.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
 PI Ffrench-Constant RH, Guo L, Hey TH, Merlo DJ, Orr GL;
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
 PI Strickland JA, Sukhapinda K;
 XX
 XX WPT: 1998-179427/16.
 DR N-PSDB; AAV29928.
 XX
 PT Isolated toxins from *Photobhabdus luminescens* strains - useful for
 PT control of insect pests
 XX
 XX Claim 34; Pages 224-231; 321pp; English.
 XX
 CC The present sequence represents a protein named TcdA of the bacterium
 CC *Photobhabdus luminescens* (W-14). This is a symbiotic bacterium of the
 CC nematodes of the *Heterorhabditis* genus. The bacterium has at least 4
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
 CC produced from these regions that are associated with insecticidal
 CC activity. The native toxins are secreted proteins. The proteins are
 CC toxic to insects upon exposure and especially when ingested. The
 CC nucleic acid sequence can be used to produce transgenic plants,
 CC baculoviruses or microbial hosts for toxin production. They can be used
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
 CC codling moth, corn earworm, European corn borer or tobacco hornworm
 CC or budworm.
 XX
 SQ Sequence 2516 AA;
 Alignment Scores:
 Pred. No.: 9,52e-46 Length: 2516
 Score: 434.50 Matches: 87
 Percent Similarity: 72.30% Conservative: 20
 Best Local Similarity: 58.78% Mismatches: 38
 Query Match: 54.79% Indels: 3
 DB: 19 Gaps: 1
 US-09-856-221-1 (1-449) x AAW56572 (1-2516)
 QY 13 CACCTTATTGGCGGATAACCTTATTTTCATTGATAACGATTGGTCAGAACCCCGTTTA 72
 Db 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907
 QY 73 GAAGAACGCCCGCAGTCMAACCATTCGGTGATCATTTATCAGCATATAAATGCGCGCACTGGCT 132
 Db 1908 AspArgAlaAlaAspLleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927
 QY 133 CAACGCGCGCCCTTGGCGCGGAAA-----CGTACTGCAAAATTCGTTAAACCGCTTTG 183
 Db 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947
 QY 184 TTCCTCTCCTCAGATAAACAACAAATTCGTAAGTACTGGCAGACGTTAGCACACGCCTA 243
 Db 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967
 QY 244 TATTAACCTAGCTATATCTGACATATGATGGTCAGCCGCTTGTCTATCCCATCTATGGC 303
 Db 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987
 QY 304 ACACGACGACATCCGTCCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 363
 Db 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007
 QY 364 GATTTGCTCGGACACTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 423
 Db 2008 LysLeuProGlnSerPheMetSerLeuIlePargPheProHisMetLeuGluAsnAlaArg 2027
 QY 424 TGGGAGTACGCCAACCACTGATACAG 447

Db 2028 GlyMetValSerGlnLeuThrGln 2035
 RESULT 3
 AAB72609
 ID AAB72609 standard; Protein; 2516 AA.
 XX
 XX AAB72609;
 XX
 XX 04-MAY-2001 (first entry)
 XX
 XX Photobhabdus tcdA toxin.
 XX
 XX TcdA; TcdA; insect toxin; plant; insect resistance.
 XX
 XX Photobhabdus sp.
 OS
 PN WO200111029-A1.
 XX
 XX 15-FEB-2001.
 XX
 XX 11-AUG-2000; 2000WO-US22237.
 PF
 PR 11-AUG-1999; 99US-0148356.
 XX
 XX (DOWC) DOW AGROSCIENCES LLC.
 PA
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 XX
 XX WPI: 2001-191536/19.
 DR N-PSDB; AAF58778.
 XX
 XX Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm
 XX
 PS Disclosure; Page 51-61; 106pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC modified versions of the *Photobhabdus* TcdA and TcdB toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon expression in plants, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC *Photobhabdus* TcdA protein.
 XX
 SQ Sequence 2516 AA;
 Alignment Scores:
 Pred. No.: 9,52e-46 Length: 2516
 Score: 434.50 Matches: 87
 Percent Similarity: 72.30% Conservative: 20
 Best Local Similarity: 58.78% Mismatches: 38
 Query Match: 54.79% Indels: 3
 DB: 22 Gaps: 1
 US-09-856-221-1 (1-449) x AAB72609 (1-2516)
 QY 13 CACCTTATTGGCGGATAACCTTATTTTCATTGATAACGATTGGTCAGAACCCCGTTTA 72
 Db 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907
 QY 73 GAAGAACGCCCGCAGTCMAACCATTCGGTGATCATTTATCAGCATATAAATGCGCGCACTGGCT 132
 Db 1908 AspArgAlaAlaAspLleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927
 QY 133 CAACGCGCGCCCTTGGCGCGGAAA-----CGTACTGCAAAATTCGTTAAACCGCTTTG 183
 Db 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947
 QY 184 TTCCTCTCCTCAGATAAACAACAAATTCGTAAGTACTGGCAGACGTTAGCACACGCCTA 243
 Db 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967

QY 244 TATACTTAGTCATATCTGACATTTGATGGTCAGCGCTTGTCTATTACCCATCTATGCG 303
 Db 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987
 QY 304 ACACCCAGCAGATCCGTCCTACTGCTTAGTCTGCTGCGCTCACCGCTCACAGGGGAGG 363
 Db 1988 ThrProAlaAspProLysAlaLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007
 QY 364 GATTTCCTCGGACAGTAATCCGATGATCCGTTTCCGATTATCTCGAAATGCGCAAG 423
 Db 2008 LysLeuProGlnSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2027
 QY 424 TGGGGAGTGACCACTGATACAG 447
 Db 2028 GlyMetValSerGlnLeuThrGln 2035
 RESULT 4
 ID AAB72611 standard; Protein; 2517 AA.
 XX
 AC AAB72611;
 DT 04-MAY-2001 (first entry)
 XX
 DE Modified Photorhabdus tcda toxin.
 KW Tcda; Tcda; insect toxin; plant; insect resistance.
 XX
 OS Photorhabdus sp.
 OS Synthetic.
 XX
 PN WO200111029-A1.
 PD 15-FEB-2001.
 XX
 PF 11-AUG-2000; 2000WO-US22237.
 XX
 PR 11-AUG-1999; 99US-0148356.
 XX
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 DR WPI: 2001-191536/19.
 DR N-PSDB; AAF58780.
 XX
 PT Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm -
 XX
 PS Claim 1; Page 72-83; 106pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC modified versions of the Photorhabdus Tcda and Tcda toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC modified Photorhabdus Tcda protein.
 XX
 SQ Sequence 2517 AA:
 Alignment Scores:
 Pred. No.: 9,52e-46 Length: 2517
 Score: 434.50 Matches: 87
 Percent Similarity: 72.30% Conserves: 20
 Best Local Similarity: 58.78% Mismatches: 38
 Query Match: 54.79% Indels: 3
 DB: 22 Gaps: 1
 US-09-856-221-1 (1-449) x AAB72611 (1-2517)

QY 13 CACCTTATGGCGGATAACCTTATTTTTCATTGATAACGATTGGTCAGAACCCCGTTTA 72
 Db 1889 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1908
 QY 73 GAAGAAGCCCGCAGTCAACCATTCGTGATCATATACGATATAAAATGCGCAACTGCGT 132
 Db 1909 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1928
 QY 133 CAAGCGCGCGCTTCCGCGCGAAA-----CGTACTGCAAAATTCGTTAAACCGCTTTG 183
 Db 1929 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1948
 QY 184 TTCCTTCCTCAGATAAACAACAAAGTTACTGGCAGAGCTTAGCACAACCCCTA 243
 Db 1949 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1968
 QY 244 TATACTTAGTCATATCTGACATTTGATGGTCAGCGCTTGTCTATTACCCATCTATGCG 303
 Db 1969 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1988
 QY 304 ACACCCAGCAGATCCGTCCTACTGCTTAGTCTGCGTCACCGCTCACAGGGGAGG 363
 Db 1989 ThrProAlaAspProLysAlaLeuSerAlaAlaValAlaThrSerGlnGlyGly 2008
 QY 364 GATTTCCTCGGACAGTAATCCGATGATCCGTTTCCGATTATCTCGAAATGCGCAAG 423
 Db 2009 LysLeuProGlnSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2028
 QY 424 TGGGGAGTGACCACTGATACAG 447
 Db 2028 GlyMetValSerGlnLeuThrGln 2036
 RESULT 5
 ID AAB72614 standard; Protein; 2537 AA.
 XX
 AC AAB72614;
 DT 04-MAY-2001 (first entry)
 XX
 DE Tcda toxin-zeln ER signal peptide fusion protein.
 KW Tcda; Tcda; insect toxin; plant; insect resistance.
 XX
 OS Chimeric - Photorhabdus sp.
 OS Chimeric - Zea mays.
 PN WO200111029-A1.
 PD 15-FEB-2001.
 XX
 PF 11-AUG-2000; 2000WO-US22237.
 XX
 PR 11-AUG-1999; 99US-0148356.
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 DR WPI: 2001-191536/19.
 DR N-PSDB; AAF58783.
 XX
 PT Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially
 PT corn rootworm -
 XX
 PS Example 1; Page 94-104; 106pp; English.
 CC The present invention provides the protein and coding sequences of
 CC modified versions of the Photorhabdus Tcda and Tcda toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce

CC transgenic plants with insect resistance. The present sequence is a
 CC fusion protein comprising the Photobhabdus tcdA toxin and the maize zein
 CC protein ER signal peptide.

XX SQ Sequence 2537 AA;

Alignment Scores:

Pred. No.: 9,55e-46 Length: 2537
 Score: 434.50 Matches: 87
 Percent Similarity: 72.30% Conservative: 20
 Best Local Similarity: 58.78% Mismatches: 38
 Query Match: 54.79% Indels: 3
 DB: 22 Gaps: 1

US-09-856-221-1 (1-449) x AAB72614 (1-2537)

QY 13 CACCTTATGGCGATACCTTATTTTCATTGGATACGATTGGTCAGAACCCGTTTA 72
 Db 1909 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1928
 QY 73 GAAGAAGCCCGCAGTCAAAACCATTCGTGATCATTTATCAGCATATAAAATGCGGCAACTGGCT 132
 Db 1929 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1948
 QY 133 CAACGCGCGCTTCGCGCGCAAA-----CGTACTGCAATTCGTAAACCGCTTGG 183
 Db 1949 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1968
 QY 184 TTCCTCTCAGATAACAAACAAAGTTACTGGCAGAGCTTAGCACACCGCTTA 243
 Db 1969 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1988
 QY 244 TATAACTTACGTACATAAATGACAAATGATGGTCAGCGCTGTGCATACCATCTATGG 303
 Db 1989 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 2008
 QY 304 ACACGACGAGATCCGCTGCTACTGCTTAGTGGCGCTCAGCGCTCACAAGCGGAGG 363
 Db 2009 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2028
 QY 364 GATTGCTCGGACAGTAAATCCGATACCGATTCGATTTTCCGATTTATCTCGAAATGCCAAG 423
 Db 2029 LysLeuProGluSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2048
 QY 424 TGGGAGTGACCACTGATACAG 447
 Db 2049 GlyMetValSerGlnLeuThrGln 2056

RESULT 6

AAY33729
 ID AAY33729 standard; Protein: 2522 AA.

XX AC AAY33729;

XX DT 09-NOV-1999 (first entry)

XX DE Photobhabdus luminescens hph2-encoded insecticidal toxin.

XX KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX OS Photobhabdus luminescens.

XX XX W09942589-A2.

XX PN 26-AUG-1999.

XX PF 18-FEB-1999; 99WO-EP01015.

XX PR 20-JAN-1999; 99US-0116439.

XX PR 20-FEB-1998; 98US-0027080.

XX XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX

PI Anderson AR, Chen JS, Dunn MM, Hart HP, Kramer VC;
 PI Morgan MK, Warren GW;

XX WPI; 1999-527479/44.

DR N-PSDB; AAZ06831.

XX New nucleic acid from Photobhabdus luminescens encoding insecticidal
 XX toxins, used for making resistant transgenic plants

XX Claim 26; Page 133-140; 148pp; English.

CC This sequence represents the hph2 gene encoded insecticidal
 CC toxin from Photobhabdus luminescens. It is one of three insecticidal

CC toxins (AAY33728-Y33730) encoded by open reading frames (orfs) in a 38kb
 CC fragment of P. luminescens DNA (AAZ06831). The hph2 gene was identified
 CC using probe f2 which was amplified from P. luminescens genomic DNA using
 CC PCR primers AAZ06829-206830. P. luminescens is a member of the
 CC Enterobacteriaceae family and is a symbiotic bacterium of the
 CC nematodes of the genus Heterorhabditis. The nematodes colonise
 CC insect larvae, kill them, and their offspring feed on the dead
 CC larvae. However, the insecticidal agents are produced by P.
 CC luminescens rather than the nematodes. The toxins have activity against
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni),
 CC European Corn Borer (Ostrinia nubilalis) and Fall Armyworm
 CC (Spodoptera frugiperda) and also against Coleopteran insects
 CC (e.g., Colorado Potato Beetle, Leptinotarsa decimlineata). In
 CC addition the toxins are active against strains resistant to known
 CC insecticides. The DNA sequence can be used to generate transgenic plants
 CC of various species that are resistant to economically important insect
 CC pests and also for recombinant production of the toxins for use as
 CC insecticides.

XX SQ Sequence 2522 AA;

Alignment Scores:

Pred. No.: 5,52e-45 Length: 2522
 Score: 428.50 Matches: 86
 Percent Similarity: 73.97% Conservative: 22
 Best Local Similarity: 58.90% Mismatches: 37
 Query Match: 54.04% Indels: 1
 DB: 20 Gaps: 1

US-09-856-221-1 (1-449) x AAY33729 (1-2522)

QY 13 CACCTTATGGCGATACCTTATTTTCATTGGATACGATTGGTCAGAACCCGTTTA 72

Db 1896 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpAsnAspProArgLeu 1915

QY 73 GAAGAAGCCCGCAGTCAAAACCATTCGTGATCATTTATCAGCATATAAAATGCGGCAACTGGCT 132

Db 1916 AspLysAlaAlaAspIleThrThrGlnSerAlaHisSerSerIleValAlaLeuArg 1935

QY 133 CAACGCGCG-----GCCTTGGCGGCAACCTACTGCAAAATTCGTTAAACCGCTTTGTCCTT 189

Db 1936 GlnSerThrProAlaLeuLeuSerLeuArgSerAlaAlaValAlaThrSerGlnGlyGlyLysLeu 1955

QY 190 CCTCAGATAACAAACAACTGCAAAAGTTACTGCGAGAGCTTAGCACAGCCCTATATAAC 249

Db 1956 ProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgValTyrAsn 1975

QY 250 TTACGTCATAATCTGACAAATGTAGTGGTCAGCGCTTGTTCATTACCCATCTATGCGACACA 309

Db 1976 LeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAlaThrPro 1995

QY 310 GCAGATCCGTCGCTACTGCTTAGTGGCTGCGCTCACAGCGCGGAGGGATTTG 369

Db 1996 AlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGlyLysLeu 2015

QY 370 CCTCGGACAGTAATCCCGATGTACCGCTTTTCGATTTATCTGGAATATCCCAAGTGGGGA 429

Db 2016 ProGluSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArgSerMet 2035

CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC codling moth, corn earworm, European corn borer or tobacco hornworm
CC or budworm.
XX Sequence 2504 AA;

Alignment Scores: 2.87e-43 Length: 2504
Pred. No.: 415.00 Matches: 90
Score: 415.00
Percent Similarity: 71.33% Conservative: 17
Best Local Similarity: 60.00% Mismatches: 41
Query Match: 52.33% Indels: 3
DB: 19 Gaps: 1

US-09-856-221-1 (1-449) x AAW56557 (1-2504)

QY 5 CAGGCACTCACCTTATTGGCGGATAA-CCTTATTTCATTGGATAACGATTGTCAGAA 63
Db 1880 GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrThrTrpAlaAsn 1899
QY 64 CCCCGTTTACAGAACGCCAGTCAAAACCACTTCGTGATCATATACGACATAAAATGCGG 123
Db 1900 ProThrLeuGlyAsnAlaAlaSerLysThrGlnGlnValArgGlnInValLeuThr 1919
QY 124 CAACCTGCGT-----CAACGCGCGCGCTTCCGCGGAAACGCTACTGCAAAATTCGTTAAC 177
Db 1920 GlnLeuArgLeuAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1939
QY 178 GCTTTGTTCTCTCCTCAGATAAACAACAACTTACTGCGACAGCTTAGCACAA 237
Db 1940 AlaLeuPheLeuProGlnGluAsnSerLysLeuGlyThrTrpArgThrLeuAlaGln 1959
QY 238 CGCCTATATACCTACCTCATAATCTGACAAATTCGTGATGCTGACCGTTCATTACCCATC 297
Db 1960 ArgMetPheAsnLeuArgHisAsnLeuSerLysLeuGlyGlnProLeuSerLeuProLeu 1979
QY 298 TATCGGACACAGAGATCCGCTACTGCTTAGTGTGCTGCGTCCGCTCAGCGCTCACAAAGC 357
Db 1980 TyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly 1999
QY 358 GGAGGGGATTGCTCGGACAGTAAATGCGGATGCTACCGTTTCCTGATTATCTGGAAAT 417
Db 2000 GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGly 2019
QY 418 GCCAAGTGGGAGTACCCCACTGATACAG 447
Db 2020 AlaArgGlyLeuValAsnGlnLeuIleGln 2029

RESULT 9
AAB72610
ID AAB72610 standard; Protein: 2504 AA.
XX AC AAB72610;
XX XX

04-MAY-2001 (first entry)

Photorhabdus tcba toxin.

TcdA; TcdB; insect toxin; plant; insect resistance.

Photorhabdus sp.

WO200111029-A1.

15-FEB-2001.

11-AUG-2000; 2000WO-US22237.

11-AUG-1999; 99US-0148356.

PA (DOWC) DOW AGROSCIENCES LLC.
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
PI Sukhapinda K, Merlo AO;
XX WPI; 2001-191536/19.
DR N-PSDB; AAF58779.
XX Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm
XX Disclosure; Page 62-72; 106pp; English.
PS The present invention provides the protein and coding sequences of
CC modified versions of the Photorhabdus TcdA and TcdB toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC Photorhabdus TcdA protein.
XX SQ Sequence 2504 AA;

Alignment Scores: 2.87e-43 Length: 2504
Pred. No.: 415.00 Matches: 90
Score: 415.00
Percent Similarity: 71.33% Conservative: 17
Best Local Similarity: 60.00% Mismatches: 41
Query Match: 52.33% Indels: 3
DB: 22 Gaps: 1

US-09-856-221-1 (1-449) x AAB72610 (1-2504)

QY 5 CAGGCACTCACCTTATTGGCGGATAA-CCTTATTTCATTGGATAACGATTGTCAGAA 63
Db 1880 GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrThrTrpAlaAsn 1899
QY 64 CCCCGTTTACAGAACGCCAGTCAAAACCACTTCGTGATCATATACGACATAAAATGCGG 123
Db 1900 ProThrLeuGlyAsnAlaAlaSerLysThrGlnGlnValArgGlnInValLeuThr 1919
QY 124 CAACCTGCGT-----CAACGCGCGCGCTTCCGCGGAAACGCTACTGCAAAATTCGTTAAC 177
Db 1920 GlnLeuArgLeuAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1939
QY 178 GCTTTGTTCTCTCCTCAGATAAACAACAACTTACTGCGACAGCTTAGCACAA 237
Db 1940 AlaLeuPheLeuProGlnGluAsnSerLysLeuGlyThrTrpArgThrLeuAlaGln 1959
QY 238 CGCCTATATACCTACCTCATAATCTGACAAATTCGTGATGCTGACCGTTCATTACCCATC 297
Db 1960 ArgMetPheAsnLeuArgHisAsnLeuSerLysLeuGlyGlnProLeuSerLeuProLeu 1979
QY 298 TATCGGACACAGAGATCCGCTACTGCTTAGTGTGCTGCGTCCGCTCAGCGCTCACAAAGC 357
Db 1980 TyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly 1999
QY 358 GGAGGGGATTGCTCGGACAGTAAATGCGGATGCTACCGTTTCCTGATTATCTGGAAAT 417
Db 2000 GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGly 2019
QY 418 GCCAAGTGGGAGTACCCCACTGATACAG 447
Db 2020 AlaArgGlyLeuValAsnGlnLeuIleGln 2029

RESULT 10
AAB72612
ID AAB72612 standard; Protein: 2505 AA.
XX AC AAB72612;
XX XX

04-MAY-2001 (first entry)

DE Modified Photorhabdus tcbA toxin.
XX TcBa; TcBa; insect toxin; plant; insect resistance.
XX Photorhabdus sp.
OS Synthetic.
XX WO20011029-A1.
XX 15-FEB-2001.
XX 11-AUG-2000; 2000WO-US22237.
XX 11-AUG-1999; 99US-0148356.
XX (DOWC) DOW AGROSCIENCES LLC.
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
PI Sukhapinda K, Merlo AO;
XX WPI; 2001-191536/19.
DR N-PSDB; AAF58781.
XX Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm -
XX Claim 1; Page 93-93; 106pp; English.
XX The present invention provides the protein and coding sequences of
CC modified versions of the Photorhabdus TcBa and TcBa toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC modified Photorhabdus TcBa protein.
XX Sequence 2505 AA;
SQ
Alignment Scores:
Pred. No.: 6,92e-43 Length: 2505
Score: 412.00 Matches: 90
Percent Similarity: 71.33% Conservatives: 17
Best Local Similarity: 60.00% Mismatches: 41
Query Match: 51.95% Indels: 3
DB: 22 Gaps: 1
US-09-856-221-1 (1-449) x AAB72612 (1-2505)
QY 5 CAGGCATCACCCTTATGGCGATAA-CCATTATTTTCATTGATACGATGTCAGAA 63
Db 1881 GlnAlaLeuAsnLeuGlyAspGluProGlnValMetLeuSerThrTrpAlaAsn 1900
QY 64 CCCCTTTAGAGAGCCCGCAGTCAACACCATTCGTGATCATATACGATATAAATCGG 123
Db 1901 ProThrLeuGlyAsnAlaAlaSerLysThrThrGlnGlnValArgGlnValLeuThr 1920
QY 124 CAACCTGGCT-----CAACGCGCGCTGTCGCGGAAACGTAAGTTCGTTAAC 177
Db 1921 GlnLeuArgGlnAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1940
QY 178 GCTTTGTCCTCTCCAGATAAACAAAACGTAAGTACTGGCAGACCTTAGCAAA 237
Db 1941 AlaLeuPheLeuProGlnGluAsnSerLysLeuLysGlyTyrTrpArgThrLeuAlaGln 1960
QY 238 CCCCTATTAACCTAGCTCATATCTGACATTCGATGTCAGCGGTGTCATTACCCATC 297
Db 1961 ArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlnProLeuSerLeuProLeu 1980
QY 298 TATGGACACAGCAGATCCGCTGCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 357
Db 1981 TyrAlaLysProAlaAspProLysAlaLeuSerAlaAlaValSerAlaSerGlnGly 2000
QY 358 GGAGGGGATTGGCTCGGACAGTAATGCGGATGACCGTTTCCGATTATTCTGGAAAT 417

Db 2001 GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGly 2020
QY 418 GCCAAGTGGGGAGTGCACCAACTGATACAG 447
Db 2021 AlaArgGlyLeuValAsnGlnLeuIleGln 2030
RESULT 11
AAW17871
ID AAW17871 standard; Protein; 2504 AA.
XX AAW17871;
XX 29-JAN-1998 (first entry)
XX Photorhabdus luminescens insect toxin protein TcBa.
XX Insecticide; insect; toxin; pest control; biological control;
KW Photorhabdus luminescens; TcBa; Southern corn rootworm;
KW Colorado potato beetle; Western corn rootworm; meal worm;
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
KW Diptera, Dictyoptera; Acarina; Homoptera.
XX Photorhabdus luminescens strain W-14 (ATCC 55397).
XX WO9717432-A1.
XX 15-MAY-1997.
XX 06-NOV-1996; 96WO-US18003.
XX 28-AUG-1996; 96US-0705484.
XX 06-NOV-1995; 95US-0007255.
XX 28-FEB-1996; 96US-0608423.
XX (WISC) WISCONSIN ALUMNI RES FOUND.
XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI Strickland JA;
XX WPI; 1997-281022/25.
DR N-PSDB; AAT68836.
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
PT insect control
PS Claim 8; Page 119-129; 276pp; English.
XX This polypeptide comprises a specifically claimed 280.6 kDa
CC insecticidal toxin, TcBa, of Photorhabdus luminescens that is
CC proteolytically cleaved into 207.6 kDa (see AAW18302) and 62.9 kDa
CC polypeptides (see AAW18303). The TcBa polypeptide can be expressed
CC in host cells using a gene (see AAT68836) isolated from a genomic
CC library. Claimed toxin proteins of P. luminescens (see AAW17871,
CC AAW17884-89, AAW17899-900, AAW18301-06) can be applied to, or
CC genetically engineered into, insect larvae food and plants for insect
CC control. The Photorhabdus toxins are particularly effective against
CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,
CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black
CC cutworm, cabbage looper, codling moth, corn earworm, European corn
CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and
CC are also active against insects of the orders Hymenoptera, Diptera
CC Dictyoptera, Acarina and Homoptera. (All claimed).
XX Sequence 2504 AA;
SQ
Alignment Scores:
Pred. No.: 1.24e-42 Length: 2504

Alignment Scores:
Pred. No.: 1.24e-35 Length: 573
Score: 353.00 Matches: 68
Percent Similarity: 84.38% Conservative: 13
Best Local Similarity: 70.83% Mismatches: 15
Query Match: 44.51% Indels: 0
DB: 18 Gaps: 0

US-09-856-221-1 (1-449) x AAW18303 (1-573)

QY 160 ACTGCAAAATCGTTACCGCTTTCTCTCAGATAAACAATAAAGTCAAGTTAC 219
DB 3 ThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLysLeuLysGlyTyr 22
QY 220 TGGCAGACGCTAGCAGCGCTATATACCTAGCTATATCTGACAAATGATGTCAG 279
DB 23 TrpArgThrLeuAlaGlnArgMetPheAsnLeuArgHisAsnLeuSerLysLeuAspGlyGln 42
QY 280 CCCTGTCTATTACCATCTATGCGACACACGAGATCCGCTCTGCTTACTGCTGCTGCC 339
DB 43 ProLeuSerLeuProLeuTyrAlaLysProAlaAspProLysAlaLeuSerAlaAla 62
QY 340 GTACACGCTCAGACGCGAGGAGTTCCTCGGACAGATGCGGATGATGCTGCTTTT 399
DB 63 ValSerAlaSerGlnGlyAlaAspLeuProLysAlaProLeuThrLeuHisArgPhe 82
QY 400 CCGATTATTCTGGAATGCCAATGGGAGTGACCCCACTGATACAG 447
DB 83 ProGlnMetLeuGluGlyAlaArgGlyLeuValAsnGlnLeuIleGln 98

RESULT 15

AAW56559
ID AAW56559 standard; Protein: 573 AA.

AC AAW56559;

XX 07-AUG-1998 (first entry)

DE Toxin TcbAIII, encoded by the tcbA gene from genomic region tcb.

XX Photorhabdus luminescens W-14; nematode; symbiotic;
KW Heterorhabdus; tcb; tcb; tcb; insecticidal activity; toxin;
KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW Tobacco hornworm; budworm.

XX Photorhabdus luminescens.

XX W09808932-A1.

XX 05-MAR-1998.

XX 05-MAY-1997; 97WO-US07657.

XX 06-NOV-1996; 96WO-US18003.

XX 28-AUG-1996; 96US-0705484.

XX 06-NOV-1996; 96US-0743699.

XX (DOWC) DOWELANCO.

PA (WISC) WISCONSIN ALUMNI RES FOUND.

XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI Strickland JA, Sukhapinda K;

XX WPI; 1998-179427/16.

XX N-PSDB; AAV29987.

XX Isolated toxins from Photorhabdus luminescens strains - useful for

PT control of insect pests

XX

PS Claim 34; Pages 261-263; 321pp; English.

XX The present sequence represents a protein named TcbAIII of the bacterium
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
CC nematodes of the Heterorhabdus genus. The bacterium has at least 4
CC distinct genomic regions, tcb, tcb, tcb, and tcb. Peptide products are
CC produced from these regions that are associated with insecticidal
CC activity. The native toxins are secreted proteins. The proteins are
CC toxic to insects upon exposure and especially when ingested. The
CC nucleic acid sequence can be used to produce transgenic plants,
CC baculoviruses or microbial hosts for toxin production. They can be used
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC codling moth, corn earworm, European corn borer or tobacco hornworm
CC or budworm.

XX Sequence 573 AA;

QY Alignment Scores:

Pred. No.: 1.24e-35 Length: 573
Score: 353.00 Matches: 68
Percent Similarity: 84.38% Conservative: 13
Best Local Similarity: 70.83% Mismatches: 15
Query Match: 44.51% Indels: 0
DB: 19 Gaps: 0

US-09-856-221-1 (1-449) x AAW56559 (1-573)

QY 160 ACTGCAAAATCGTTACCGCTTTCTCTCAGATAAACAATAAAGTCAAGTTAC 219

DB 3 ThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLysLeuLysGlyTyr 22

QY 220 TGGCAGACGCTAGCAGCGCTATATACCTAGCTATATCTGACAAATGATGTCAG 279

DB 23 TrpArgThrLeuAlaGlnArgMetPheAsnLeuArgHisAsnLeuSerLysLeuAspGlyGln 42

QY 280 CCCTGTCTATTACCATCTATGCGACACACGAGATCCGCTCTGCTTACTGCTGCTGCC 339

DB 43 ProLeuSerLeuProLeuTyrAlaLysProAlaAspProLysAlaLeuSerAlaAla 62

QY 340 GTACACGCTCAGACGCGAGGAGTTCCTCGGACAGATGCGGATGATGCTGCTTTT 399

DB 63 ValSerAlaSerGlnGlyAlaAspLeuProLysAlaProLeuThrLeuHisArgPhe 82

QY 400 CCGATTATTCTGGAATGCCAATGGGAGTGACCCCACTGATACAG 447

DB 83 ProGlnMetLeuGluGlyAlaArgGlyLeuValAsnGlnLeuIleGln 98

Search completed: January 15, 2003, 15:50:23

Job time : 53.5673 secs



2

3

4

5

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 15, 2003, 15:45:23 ; Search time 14.4662 Seconds
(without alignments)
1826.456 Million cell updates/sec

Title: US-09-856-221-1
Perfect score: 793
Sequence: 1 tttgcaggcactcaccttat.....gtgaccccaactgatacagtt 449

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q/-cgn2_1/USPTO.spool/US09856221/runat_15012003_153926_12521/app.query.fasta_1.1948
-DB-Issued_Patents_AA -OPMT-fastan -SUFFIX-rai -MINMATCH-0.1 -LOOPCL-0
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST-45 -DOCALIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15
-MODE-LOCAL -OUTPMT-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-USER-US09856221@cgn.1.1.24 @runat_15012003_153926_12521 -NCPU-6 -ICPU-3
-NO_XLPVY -NO_MAP -LARGEQUERY -NEG_SCORES-0 -WAIT -LONGLOG -DEV_TIMEOUT-120
-WARN_TIMEOUT-30 -THREADS-1 -XGAPOPT-10 -XGAPEXT-0.5 -FCAPOPT-6 -FCAPEXT-7
-YGAPOPT-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/6C-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	428.5	54.0	2522	4	US-09-251-645-13
2	71.5	9.0	942	3	US-09-074-579-1
3	71.5	9.0	942	4	US-09-388-774-1
4	71.5	9.0	2441	1	US-08-194-468-2
5	71.5	9.0	2441	3	US-08-961-739-2
6	71.5	9.0	2441	4	US-09-514-247A-8
7	70	8.8	216	2	US-08-840-683-9
8	70	8.8	216	2	US-08-555-722-9
9	70	8.8	216	4	US-09-384-301-9
10	70	8.8	983	4	US-09-134-001C-3814
11	69.5	8.8	342	3	US-08-978-456-2
12	69.5	8.8	342	4	US-09-369-700-2
					Sequence 13, Appl
					Sequence 1, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 8, Appl
					Sequence 9, Appl
					Sequence 9, Appl
					Sequence 9, Appl
					Sequence 3814, Ap
					Sequence 2, Appl

13	69	8.7	201	1	US-08-444-083-8	Sequence 8, Appl
14	69	8.7	201	1	US-08-286-304-8	Sequence 8, Appl
15	69	8.7	201	1	US-08-442-745-8	Sequence 8, Appl
16	69	8.7	201	1	US-08-443-129-8	Sequence 8, Appl
17	69	8.7	201	1	US-08-443-952-8	Sequence 8, Appl
18	69	8.7	201	1	US-08-443-130-8	Sequence 8, Appl
19	69	8.7	201	1	US-08-792-019B-11	Sequence 11, Appl
20	69	8.7	201	3	US-09-106-182-4	Sequence 4, Appl
21	69	8.7	201	3	US-08-988-819-11	Sequence 11, Appl
22	69	8.7	201	3	US-08-988-911-8	Sequence 8, Appl
23	69	8.7	201	4	US-09-016-534-11	Sequence 11, Appl
24	69	8.7	201	5	PCT-US95-04467-8	Sequence 8, Appl
25	68	8.6	448	3	US-08-476-509B-2	Sequence 2, Appl
26	68	8.6	486	3	US-08-348-518C-2	Sequence 2, Appl
27	68	8.6	1274	4	US-09-095-443-2	Sequence 2, Appl
28	67.5	8.5	834	3	US-08-539-205A-6	Sequence 6, Appl
c 29	67.5	8.6	837	1	US-07-923-976-2	Sequence 2, Appl
c 30	66.5	8.5	951	1	US-08-162-809-2	Sequence 2, Appl
c 31	66	8.3	698	4	US-09-134-001C-3632	Sequence 3, Appl
c 32	66	8.3	1349	3	US-08-938-291A-6	Sequence 6, Appl
c 33	65	8.3	299	4	US-09-370-976-2	Sequence 2, Appl
34	65	8.2	660	4	US-09-462-606-57	Sequence 57, Appl
35	65	8.2	1417	2	US-08-559-303B-78	Sequence 78, Appl
36	65	8.2	1417	3	US-08-781-891-78	Sequence 78, Appl
37	65	8.2	1417	4	US-09-175-828-78	Sequence 78, Appl
38	64.5	8.1	506	4	US-08-942-012B-26	Sequence 26, Appl
39	64.5	8.1	624	2	US-08-879-561-3	Sequence 3, Appl
40	64.5	8.1	659	1	US-08-240-049B-16	Sequence 16, Appl
41	64.5	8.1	659	1	US-08-259-148A-20	Sequence 20, Appl
42	64.5	8.1	659	1	US-08-484-054-20	Sequence 20, Appl
43	64.5	8.1	659	2	US-07-876-941A-20	Sequence 20, Appl
44	64.5	8.1	659	4	US-08-477-292-14	Sequence 14, Appl
45	64.5	8.1	659	4	US-07-870-985A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-251-645-13
; Sequence 13, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-13

Alignment Scores:
Pred. No.: 1.2e-46 Length: 2522
Score: 428.50 Matches: 86
Percent Similarity: 73.97% Conservativeness: 22
Best Local Similarity: 58.90% Mismatches: 37
Query Match: 54.04% Indels: 1
DB: 4 Gaps: 1

US-09-856-221-1 (1-449) x US-09-251-645-13 (1-2522)

QY 13 CACCTTATTGGCGGATACCTTTATTTTCATTGGTACGATTGGTCAGAACCCGTTTA 72

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRSNOT02
; CLONE: 688183
US-09-074-579-1

Alignment Scores:
Pred. No.: 1.85          Length: 942
Score: 71.50           Matches: 27
Percent Similarity: 45.54%  Conservative: 24
Best Local Similarity: 24.11%  Mismatches: 38
Query Match: 9.02%          Indels: 23
DB: 3                    Gaps: 4

US-09-856-221-1 (1-449) x US-09-074-579-1 (1-942)
Qy 58 TCAGAACCCCGTTTGAAGAGAGCCGCGCAGTCAACACCATTCGTGATCATTTATCAGCATAAA 117
Db 118 SerGlyAspArgValGlyGluLysArgAsnLysThrGluGluAsnGlyGluLysGly 137
Qy 118 ATGGCGCAACTCGTCAACGCGCGCTTCGGCGGCAACACTGCTACTGCAAAATTCGTTAAC 177
Db 138 ThrGluIlePheArgAlaSerAlaValIleProSerLysAspLysAla----- 153
Qy 178 GCTTTGTTCCTCCTCAGATATAACAAAAAACTGCAAAAGTTACTGGCAGACGTTAGCACAA 237
Db 154 AlaPhePheLeu-----SerTyrGluGluLeuLeuGlnArg 165
Qy 238 CGCTATATAACTTACGTCAATATCTGCAATGATGATGATGATGATGATGATGATGATGATG 297
Db 166 ArgLeuGlyLysTyrGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185
Qy 298 TATCGCACACACGATCGCTGCTACTGCTTAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 357
Db 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201
Qy 358 GGAGGGGATTCGCTCGGACAGTAAATCGCGATGATC 393
Db 202 -----ValLeuProLeuHis 206
```

```

RESULT 3
US-09-388-774-1
; Sequence 1, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,579
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PP-0505 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 amino acids
; TYPE: amino acid
```



```
;
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRSN02
; CLONE: 688183
;
US-09-388-774-1
Alignment Scores:
Pred. No.: 1.85 Length: 942
Score: 71.50 Matches: 27
Percent Similarity: 45.54% Conservative: 24
Best Local Similarity: 24.11% Mismatches: 38
Query Match: 9.02% Indels: 23
DB: 4 Gaps: 4

US-09-856-221-1 (1-449) x US-09-388-774-1 (1-942)
QY 58 TCAGAACCCGTTTGAAGAGCGCGCAGTCAACCATTCGTGATCATTTATCAGCATAAA 117
DB 118 SerGlyAspArgValLysGluLysArgAsnLysThrThrGluGluAsnGlyGluLysGly 137
QY 118 ATCGCGCAACTCGTCAACGCGCGCTTCCGCGGAGAACACTCTGCAAAATTCGTTAAC 177
DB 138 ThrGluLeuPheArgAlaSerAlaValIleProSerLysAspLysAla----- 153
QY 178 GCTTTGTCCTTCCTCAGATTAACAAACAAAGTCTACTGGCAGACGTTAGCACAA 237
DB 154 AlaPhePheLeu-----SerTyrGluGluLeuLeuGlnArg 165
QY 238 CGCTATATAACTACGTCAATCTGACAAATGATGTCAGCGGTGTGCTATTACCCATC 297
DB 166 ArgLeuGlyLysTyrGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185
QY 298 TATGCGACACGACGACGTCGCTGCTAGTCTAGTGTGCTGCGGTCACCGCTCACAA 357
DB 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201
QY 358 GGAGGGGATTTCCTCGGACAGTAATGCGGATGCTAC 393
DB 202 -----ValLeuProLeuHis 206

RESULT 4
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; RESPONSIVE GENES
; NUMBER OF INVENTIONS: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-194-468-2
Alignment Scores:
Pred. No.: 3.26 Length: 2441
Score: 71.00 Matches: 34
Percent Similarity: 38.97% Conservative: 19
Best Local Similarity: 25.00% Mismatches: 45
Query Match: 8.95% Indels: 38
DB: 1 Gaps: 7

US-09-856-221-1 (1-449) x US-08-194-468-2 (1-2441)
QY 76 GAAGCGCGCAGTCAACCATTCGTGATCATTTATCAGCATAAAATCGCGCAA----- 126
DB 593 GluHisValThrGlnAspLeuArgSerHisLeuValHisLysLeuValGlnAlaIlePhe 612
QY 127 ---CTCGTCAACGCGCGCTTCCGCGGAGAACGCTACTGCAAAATTCGTTAACCGTTG 183
DB 613 ProThrProAspProAlaAlaLeuLysAspArgMetGluAsn----- 627
QY 184 TTCCTTCTCAGATAACAAAAAATCGCAA----- 213
DB 628 ---LeuValAlaTyrAlaLysLysValGluGlyAspMetTyrGluSerAlaAsnSerArg 646
QY 214 ---AGTTACTGTCAGACGCTAGCAACGCTATATACTTACGT----- 255
DB 647 AspGluTyrTyrHisLeuLeuAlaGluLysIleTyrLysIleGlnLysGluLeuGlu 666
QY 256 -----CATATCTGACAAATGATGTCAGCGGTGCTCATTTACCATC 297
DB 667 LysArgArgThrArgLeuHisLysGlnGlyIleLeuGlyAsnGlnProAlaLeuPro--- 685
QY 298 TATGCGACACGACGAGATCCGTCGCTAGTCTAGTGTGCTGCTACCGCTCACCGCTCACAA 357
DB 686 ---AlaSerGlyAlaGlnProProVal---IleProProAlaGlnSerValArgProPro 703
QY 358 GGAGGGGATTTCCTCGGACAGTAATGCGGATGCTACCGTTTTCGATT 405
DB 704 AsnGlyProLeuPro-----LeuProValAsnArgMetGlnVal 716

RESULT 5
US-08-961-739-2
; Sequence 2, Application US/08961739A
; Patent No. 6063583
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
; FILE REFERENCE: SALK1650-1
; CURRENT APPLICATION NUMBER: US/08/961,739A
; CURRENT FILING DATE: 1997-10-31
; EARLIER APPLICATION NUMBER: US 194,468
; EARLIER FILING DATE: 1994-02-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2441
```

TYPE: PRT
ORGANISM: Mus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)....(2441)
OTHER INFORMATION: Xaa = Any Amino Acid

US-08-961-739-2

Alignment Scores:

Pred. No.: 3.26 Length: 2441
Score: 71.00 Matches: 34
Percent Similarity: 38.97% Conservative: 19
Best Local Similarity: 25.00% Mismatches: 45
Query Match: 8.95% Indels: 38
DB: Gaps: 7

US-09-856-221-1 (1-449) x US-08-961-739-2 (1-2441)

QY 76 GAAGCGCCAGTCAAAACCATTCGTGATCATATACGATAAAATGCGGCAA----- 126
Db 593 GluHisValThrGlnAspLeuArgSerHisLeuValHisLysLeuValGlnAlaIlePhe 612
QY 127 ---CTCGCTCAACGCGCGCTTCGCGGGAACGCTACTGCAAAATTCGTTAACCGCTTTG 183
Db 613 ProThrProAspProAlaAlaLeuLysAspArgMetGluAsn----- 627
QY 184 TTCCTTCCTCAGATAAACAACAACTGCAA----- 213
Db 628 ---LeuValAlaTyrAlaLysValGluGlyAspMetTyrGluSerAlaAsnSerArg 646
QY 214 ---AGTTACTGCGCAGCGCTTAGCACACGCTATATACCTTACGT----- 255
Db 647 AspGluTyrTyrHisLeuLeuAlaGluLysIleTyrLysIleGlnLysGluLeuGlu 666
QY 256 -----CATATCTGACAATTCGATGTCGACCGCTGTCATTACCCATC 297
Db 667 LysArgArgThrArgLeuHisLysGlnGlyLeuGlyAsnGlnProAlaLeuPro--- 685
QY 298 TAGCGCACACCGACAGATCGCTGCTAGTCTGCTGCGGTACCGCTCACAAGGC 357
Db 686 ---AlaSerGlyAlaGlnProProVal---IleProProAlaGlnSerValArgProPro 703
QY 358 GGAGGGGATTCCTCGGACAGTAATCGCGCTACCGCTTTCCGATT 405
Db 704 AsnGlyProLeuPro-----LeuProValAsnArgMetGlnVal 716

RESULT 6

US-09-514-247A-8
; Sequence 8, Application US/09514247A
; Patent No. 6365361
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, TOMOYASU
; APPLICANT: TANIGUCHI, JUNKO
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/09/514,247A
; PRIOR FILING DATE: 2000-02-28
; PRIOR FILING DATE: 1998-08-24
; PRIOR FILING DATE: JF231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-09-514-247A-8

Alignment Scores: 3.26 Length: 2441
Pred. No.: 71.00 Matches: 34
Score:

Percent Similarity: 38.97% Conservative: 19
Best Local Similarity: 25.00% Mismatches: 45
Query Match: 8.95% Indels: 38
DB: Gaps: 7

US-09-856-221-1 (1-449) x US-09-514-247A-8 (1-2441)

QY 76 GAAGCGCCAGTCAAAACCATTCGTGATCATATACGATAAAATGCGGCAA----- 126
Db 593 GluHisValThrGlnAspLeuArgSerHisLeuValHisLysLeuValGlnAlaIlePhe 612
QY 127 ---CTCGCTCAACGCGCGCTTCGCGGGAACGCTACTGCAAAATTCGTTAACCGCTTTG 183
Db 613 ProThrProAspProAlaAlaLeuLysAspArgMetGluAsn----- 627
QY 184 TTCCTTCCTCAGATAAACAACAACTGCAA----- 213
Db 628 ---LeuValAlaTyrAlaLysValGluGlyAspMetTyrGluSerAlaAsnSerArg 646
QY 214 ---AGTTACTGCGCAGCGCTTAGCACACGCTATATACCTTACGT----- 255
Db 647 AspGluTyrTyrHisLeuLeuAlaGluLysIleTyrLysIleGlnLysGluLeuGlu 666
QY 256 -----CATATCTGACAATTCGATGTCGACCGCTGTCATTACCCATC 297
Db 667 LysArgArgThrArgLeuHisLysGlnGlyLeuGlyAsnGlnProAlaLeuPro--- 685
QY 298 TAGCGCACACCGACAGATCGCTGCTAGTCTGCTGCGGTACCGCTCACAAGGC 357
Db 686 ---AlaSerGlyAlaGlnProProVal---IleProProAlaGlnSerValArgProPro 703
QY 358 GGAGGGGATTCCTCGGACAGTAATCGCGCTACCGCTTTCCGATT 405
Db 704 AsnGlyProLeuPro-----LeuProValAsnArgMetGlnVal 716

RESULT 7

US-08-840-683-9
; Sequence 9, Application US/08840683
; Patent No. 5821051
; GENERAL INFORMATION:
; APPLICANT: ANDROPHY, ELLIOT J.
; APPLICANT: CHEN, JASON J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (Text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 216 amino acids

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
US-08-840-683-9
Alignment Scores:
Pred. No.: 1.51 Length: 216
Score: 70.00 Matches: 25
Percent Similarity: 43.59% Conservative: 9
Best Local Similarity: 32.05% Mismatches: 34
Query Match: 8.83% Indels: 10
DB: 2 Gaps: 2
US-09-856-221-1 (1-449) x US-08-840-683-9 (1-216)
QY 151 GCGAAACGCTACTGCAAAATTCGTTAAAC-----GCTTTGTTC 186
Db 100 AlaHisArgThrSerLysLeuLeuThrGlnSerGlyProSerLeuProGlnAlaThrSer 119
QY 187 CTTCTCCTCAGATAAACAACAAAGTCTACTGGCAGAGCTTTAGCACAACGCTATAT 246
Db 120 IleProAlaIleLeuLysAlaLeuGlnAspGluTrpAspAlaValMetLeuHisSerPhe 139
QY 247 AACTTACGTCATATCTGACAAATGATGTCAGCGGTTGTCATTACCATCTATGCGACA 306
Db 140 ThrLeuArgGlnGlnLeuGlnThrThrArgGlnGluLeuSerHisAlaLeuTyGlnHis 159
QY 307 CCAGCAGATCCGCTCGTACTG-----CTTAGTGCTGCGTCACCGCTCACAA 354
Db 160 AspAlaAlaCysArgValIleAlaArgLeuThrLysGluValThrAlaAlaArg 177
RESULT 8
US-08-555-722-9
; Sequence 9, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 216 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
US-08-555-722-9
Alignment Scores:
Pred. No.: 1.51 Length: 216
Score: 70.00 Matches: 25
```

Percent Similarity: 43.59% Conservative: 9
Best Local Similarity: 32.05% Mismatches: 34
Query Match: 8.83% Indels: 10
DB: 4 Gaps: 2

US-09-856-221-1 (1-449) x US-09-384-301-9 (1-216)

QY 151 CGGAACGCTACTGCAAAATCGTTACCC-----GCTTGTTC 186
||| |||||: |||||
Db 100 AlaHisGthrSerLysLeuThrGlnSerGlyProSerLeuProGlnAlaThrSer 119
||| |||||: |||||
QY 187 CTTCCTCAGATAACAAAACGCAAAAGTTACTGGCAGACGTTAGCACACGCGCTATAT 246
: ||||| ||| ||||| ||| :
Db 120 IleProAlaIleLeuLysAlaLeuGlnAspGluTrpAspAlaValMetLeuHisSerPhe 139
: ||||| ||| ||||| ||| :
QY 247 AACTTACGTCATAATCTGACAATTCAGGTCAGCCGTTGTTCATTCACCATCTATCGGACA 306
||| ||||| ||| ||||| ||| :
Db 140 ThrLeuArgGlnGlnLeuGlnThrThrArgGlnGluLeuSerHisAlaLeuTyrglnHis 159
||| ||||| ||| ||||| ||| :
QY 307 CCAGCAGATCCGTCGCTACTG-----CTTAGTGCTGCGTCACGCGCTCACAA 354
||| ||||| ||| :
Db 160 AspAlaAlaCysArgValIleAlaArgLeuThrLysGluValThrAlaAlaArg 177
||| ||||| ||| :
RESULT 10
US-09-134-001C-3814
; Sequence 3814, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3814
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3814

Alignment Scores:
Pred. No.: 2.95 Length: 983
Score: 70.00 Matches: 32
Percent Similarity: 39.45% Conservative: 26
Best Local Similarity: 21.77% Mismatches: 39
Query Match: 8.83% Indels: 50
DB: 4 Gaps: 9

US-09-856-221-1 (1-449) x US-09-134-001C-3814 (1-983)

QY 34 TATTTTCATTGATAC-----GATTGGTCAGAACCCGTTTA 72
||| |||||: |||||
Db 816 TyrrSerLeuLysAsnArgLeuAsnAspIleAlaLysAspIleAlaSerLeuSerTyrr 835
||| |||||: |||||
QY 73 GAAGAAGCCGACGTCACAAACGATTCGTCATTCATTCAGCATATAAATCGGCAACTGCGT 132
: ||||| : |||||
Db 836 MetGlnAla-----LeuValGluGluHis-----IleLysGlnIleLys 848
: ||||| : |||||
QY 133 CAAGCGCGCGCTTGGCGCGCAAGCTACTGCAAAATTCGTTTGTTCCTTCCT 192
||| |||||
Db 849 AspLysArg-----LeuPro 853
||| |||||
QY 193 CAGATAACAAAACGTAAGTTACTGGCAGACGTTAGCACACGCTTATATACTTA 252
||| |||||: |||||
Db 854 GlnValIleAsnGluAlaValSerIlePheLysAsnGlnThrAsnGlyThrTyrrAsnMet 873
||| |||||: |||||
QY 253 -----CCTCATATCTGACAATT-----GATGGTCAG----- 279
||| |||||: |||||

Db 874 IleHisTyrrThrGluAsnHisLysIleHisValLysHisSerAsnGlyGlnValPheGlu 893
||| |||||: |||||
QY 280 CCGTTGTTCATTACCCATCTATCGCACACGAGATCCGTCGCTACTGCTTAGTGTGCTGCC 339
||| |||||: |||||
Db 894 ProValGluLeu-----SerGlnSerThrLysGluLeuTyrrValAla 908
||| |||||: |||||
QY 340 GTCACCGCTCACAAGCGGAGGAGGATTTCCTCGGACAGTAATGCCGATGTACCGTTTT 399
: ||||| : |||||
Db 909 LeuArgIleSer-----LeuIleLysValLeuLysProTyrrTyrrProphe 923
||| |||||: |||||
QY 400 CCGATTATTCTGGAATGCC 420
||| |||||: |||||
Db 924 ProValIleValAspAla 930
||| |||||: |||||
RESULT 11
US-08-978-456-2
; Sequence 2, Application US/08978456
; Patent No. 6010881
; GENERAL INFORMATION:
; APPLICANT: Palmer, Leslie M.
; TITLE OF INVENTION: No. 6010881el ribg
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,456
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50444-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-978-456-2

Alignment Scores:
Pred. No.: 2.15 Length: 342
Score: 69.50 Matches: 28
Percent Similarity: 42.86% Conservative: 11
Best Local Similarity: 30.77% Mismatches: 27
Query Match: 8.76% Indels: 25
DB: 3 Gaps: 5

US-09-856-221-1 (1-449) x US-08-978-456-2 (1-342)

QY 73 GAAGAAGCCGACGTCACAAACGATTCGTCATTCATTCAGCATATAAATCGGCAACTGCGT 132
||| |||||: |||||
Db 119 AspGluArgAlaSerGlnLeuTyrrGlnAspPheLysAlaLysGlnLeuPro 138
||| |||||: |||||
QY 133 CAA-----CGCGCGCGCTTCGCGCGCAAAACGCTACTGCAAAATTCGTTAACG 177
||| |||||: |||||

```

Db 139 GlnlleThrValLysValSerAlaSerLeuAspGlyLys----- 151
QY 178 GCTTTGTCCTCCTCAGATAAACAATAAAGTTACTGGCAGAGC----- 228
Db 152 -----GlnAlaAsnAspAsnGlyGlnSerGlnTrpIleThrAsnLysGlu 166
QY 229 TTAGCACAAAGCCTATATAACTTA-----CGTCATAATCTGACAATTGATGCTCAG--- 279
Db 167 ValLysGlnAspValTyLysLeuArgHisArgHisAspAlaValLeuThrGlyArgArg 186
QY 280 -----CCGTGTCATTACCCATCTATGCGACA 306
Db 187 ThrValGluLeuAspAspProGlnTyThrThr 197

RESULT 12
US-09-369-700-2
; Sequence 2, Application US/09369700
; Patent No. 6280735
; GENERAL INFORMATION:
; APPLICANT: Palmer, Leslie M.
; TITLE OF INVENTION: No. 6280735el ribG
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09369,700
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/978,456
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd O
; REGISTRATION NUMBER: 28,35A
; REFERENCE/DOCKET NUMBER: P50444-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-369-700-2
Alignment Scores:
Pred. No.: 2.15 Length: 342
Score: 69.50 Matches: 28
Percent Similarity: 42.86% Conservative: 11
Best Local Similarity: 30.77% Mismatches: 27
Query Match: 8.76% Indels: 25
DB: 4 Gaps: 5

US-09-856-221-1 (1-449) x US-09-369-700-2 (1-342)
QY 73 GAAGAGCCCGCAGTCAACACCATTCGTCATCATATACAGCATAAATCGGCAACTGCGT 132
Db 119 AspGluArgAlaLysGlnLeuTyArgHisPheLeuAlaLysGlnLeuPro 138
QY 133 CAA-----CGCCGCGCCTTGCAGGCAACAGTACTGCAAAATTCGTTAAAC 177
|||

```

```

Db 139 GlnlleThrValLysValSerAlaSerLeuAspGlyLys----- 151
QY 178 GCTTTGTCCTCCTCAGATAAACAATAAAGTTACTGGCAGAGC----- 228
Db 152 -----GlnAlaAsnAspAsnGlyGlnSerGlnTrpIleThrAsnLysGlu 166
QY 229 TTAGCACAAAGCCTATATAACTTA-----CGTCATAATCTGACAATTGATGCTCAG--- 279
Db 167 ValLysGlnAspValTyLysLeuArgHisArgHisAspAlaValLeuThrGlyArgArg 186
QY 280 -----CCGTGTCATTACCCATCTATGCGACA 306
Db 187 ThrValGluLeuAspAspProGlnTyThrThr 197

RESULT 13
US-08-444-083-8
; Sequence 8, Application US/08444083
; Patent No. 5571675
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,083
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894P1D5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-444-083-8
Alignment Scores:
Pred. No.: 1.98 Length: 201
Score: 69.00 Matches: 25
Percent Similarity: 38.78% Conservative: 13
Best Local Similarity: 25.51% Mismatches: 31
Query Match: 8.70% Indels: 29
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x US-08-444-083-8 (1-201)

```

QY 61 GAACCCCGTTTGAAGAAGCCGCCAGTCAAAACCATTCGTGATCATATTCAGCATATAAATG 120
 Db 10 AspProGlnThrAspSerValSer-----LeuLeuProHisLeuGluAlaLysile 27
 QY 121 CGGCAACTGGTCAACGGCGCGCCTTGCCGCGAAGAGTACTGCAAAATTCGTTAACCGCT 180
 Db 28 ArgGln-----ThrHisSerLeuAlaHis 35
 QY 181 TTGTTCTCTCCAGATAAACAATACTGCAAGTTACTGGCAGACGCTTAGCACACAGC 240
 Db 36 LeuLeuThrLysTyrAlaGluGlnLeuLeuGlnGluTyrValGln-----50
 QY 241 CTATATAACTTACGTCATAATCTGACAATTCGATGGTCAGCCGTTGTCATTACCACTAT 300
 Db 51 -----LeuGlnGlyAspProPheGlyLeuProSerPhe 61
 QY 301 CGGACACGACGATCCGTCGTACTGCTGCTGCGCTCAGCGCTCAC 352
 Db 62 SerProProArgLeuProValAlaGlyLeuSerAla-ProAlaProSerHis 78

RESULT 14
 US-08-286-304-8
 ; Sequence 8, Application US/08286304
 ; Patent No. 5571893
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Joffre
 ; APPLICANT: Chien, Kenneth
 ; APPLICANT: King, Kathleen
 ; APPLICANT: Pennica, Diane
 ; APPLICANT: Wood, William
 ; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 460 Point San Bruno Blvd
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: patin (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/286,304
 ; FILING DATE: 05-AUG-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/233609
 ; FILING DATE: 25-APR-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hasak, Janet E.
 ; REGISTRATION NUMBER: 28,616
 ; REFERENCE/DOCKET NUMBER: 894P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415/225-1896
 ; TELEFAX: 415/952-9881
 ; TELEX: 910/371-7168
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 201 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-286-304-8

Alignment Scores:
 Pred. No.: 1.98
 Score: 69.00
 Percent Similarity: 38.78%
 Best Local Similarity: 25.51%
 Query Match: 8.70%
 DB: 1
 Length: 201
 Matches: 25
 Conservatives: 31
 Mismatches: 29
 Indels: 3
 Gaps: 3

US-09-856-221-1 (1-449) x US-08-286-304-8 (1-201)
 QY 61 GAACCCCGTTTGAAGAAGCCGCCAGTCAAAACCATTCGTGATCATATTCAGCATATAAATG 120
 Db 10 AspProGlnThrAspSerValSer-----LeuLeuProHisLeuGluAlaLysile 27
 QY 121 CGGCAACTGGTCAACGGCGCGCCTTGCCGCGAAGAGTACTGCAAAATTCGTTAACCGCT 180
 Db 28 ArgGln-----ThrHisSerLeuAlaHis 35
 QY 181 TTGTTCTCTCCAGATAAACAATACTGCAAGTTACTGGCAGACGCTTAGCACACAGC 240
 Db 36 LeuLeuThrLysTyrAlaGluGlnLeuLeuGlnGluTyrValGln-----50
 QY 241 CTATATAACTTACGTCATAATCTGACAATTCGATGGTCAGCCGTTGTCATTACCACTAT 300
 Db 51 -----LeuGlnGlyAspProPheGlyLeuProSerPhe 61
 QY 301 CGGACACGACGATCCGTCGTACTGCTGCTGCGCTCAGCGCTCAC 352
 Db 62 SerProProArgLeuProValAlaGlyLeuSerAla-ProAlaProSerHis 78

RESULT 15
 US-08-442-745-8
 ; Sequence 8, Application US/08442745
 ; Patent No. 5624806
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Joffre
 ; APPLICANT: Chien, Kenneth
 ; APPLICANT: King, Kathleen
 ; APPLICANT: Pennica, Diane
 ; APPLICANT: Wood, William
 ; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 460 Point San Bruno Blvd
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: patin (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/442,745
 ; FILING DATE: 17-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/233609
 ; FILING DATE: 25-APR-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/286304
 ; FILING DATE: 05-AUG-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Torchia, Timothy E.
 ; REGISTRATION NUMBER: 36,700
 ; REFERENCE/DOCKET NUMBER: 894P1D1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415/225-8674
 ; TELEFAX: 415/952-9881
 ; TELEX: 910/371-7168
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 201 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-442-745-8

Alignment Scores:

Pred. No.:	1.98	Length:	201
Score:	69.00	Matches:	25
Percent Similarity:	38.78%	Conservative:	13
Best Local Similarity:	25.51%	Mismatches:	31
Query Match:	8.70%	Indels:	29
DB:	1	Gaps:	3

US-09-856-221-1 (1-449) x US-08-442-745-8 (1-201)

Qy	61	GAACCCGGTTTGAAGAAGCCGCGCAGTCAAAACCATTCGTGATCATTTACACATAAATG	120
Db	10	AspProGlnThrAspSerSerValSer-----LeuLeuProHisLeuGluAlaLysIle	27
Qy	121	CGGCAACTGCGTCAACGGCGCGCTTCGCGCGGAACGTACTGCAAAATTCGTTAACCGCT	180
Db	28	ArgGln-----ThrHisSerLeuAlaHis	35
Qy	181	TGTGTCCTTCCTCAGATAACAAAACATGCAAACTTACTGGCAGACGTTAGCACACACGC	240
Db	36	LeuLeuThrLysTyrAlaGluGlnLeuLeuGlnGluTyrValGln-----	50
Qy	241	CTATATAACTTACGTCATAACTCTGACAAATTTGATGTCAGCCGCTGTCATTACCCACTCAT	300
Db	51	-----LeuGlnGlyAspProPheClyLeuProSerPhe	61
Qy	301	GGGACACACGACGATCCGTCGGTACTGCTTACTGTCGCGCTCACGCGCTCAC	352
Db	62	SerProArgLeuProValAlaGlyLeuSerAla-ProAlaProSerHis	78

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:46:43 ; Search time 8.42333 Seconds

(without alignments)
2119.222 Million cell updates/sec

Title: US-09-856-221-1
Perfect score: 793
Sequence: 1 tgtcagcactcaccttat.....gtgacccaactgatacagtt 449

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09856221/runat_15012003_153926_12555/app_query.fasta_1.1948
-DB=PublishedApplications_AA -OFFT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09856221.ecgn_1.14_@runat_15012003_153926_12555
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PC1_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	434.5	54.8	2516	10	US-09-817-514A-2	Sequence 2, Appli
2	415	52.3	2504	10	US-09-817-514A-8	Sequence 8, Appli
3	76	9.6	1447	10	US-09-808-571A-2	Sequence 2, Appli
4	76	9.6	2799	9	US-10-151-736-4	Sequence 4, Appli

5	73.5	9.3	338	9	US-09-933-638A-12	Sequence 12, Appl
6	73.5	9.3	371	9	US-09-849-243-16	Sequence 16, Appl
7	71.5	9.0	694	9	US-09-944-413-55	Sequence 55, Appl
8	71.5	9.0	694	9	US-09-944-403-55	Sequence 55, Appl
9	71.5	9.0	694	9	US-09-944-896-55	Sequence 55, Appl
10	71.5	9.0	694	9	US-09-944-944-55	Sequence 55, Appl
11	71.5	9.0	694	9	US-09-944-907-55	Sequence 55, Appl
12	71.5	9.0	694	9	US-09-944-929-55	Sequence 55, Appl
13	71.5	9.0	694	9	US-10-174-590-60	Sequence 60, Appl
14	71.5	9.0	694	9	US-10-176-758-60	Sequence 60, Appl
15	71.5	9.0	694	10	US-09-866-028-55	Sequence 55, Appl
16	71.5	9.0	694	10	US-09-944-449-55	Sequence 55, Appl
17	71.5	9.0	694	10	US-09-944-457-55	Sequence 55, Appl
18	71.5	9.0	694	10	US-09-944-862-55	Sequence 55, Appl
19	71.5	9.0	694	10	US-09-945-587-55	Sequence 55, Appl
20	71.5	9.0	694	10	US-09-945-015-55	Sequence 55, Appl
21	71.5	9.0	694	10	US-09-944-396-55	Sequence 55, Appl
22	71.5	9.0	694	10	US-09-944-097-55	Sequence 55, Appl
23	71.5	9.0	694	10	US-09-944-432-55	Sequence 55, Appl
24	71.5	9.0	694	10	US-09-943-762-55	Sequence 55, Appl
25	71.5	9.0	694	10	US-09-944-654-55	Sequence 55, Appl
26	71.5	9.0	694	10	US-09-943-861A-55	Sequence 55, Appl
27	71.5	9.0	694	12	US-10-052-586-60	Sequence 60, Appl
28	71.5	9.0	942	10	US-09-828-423-1	Sequence 1, Appl
29	71	8.8	2441	12	US-10-109-886-8	Sequence 8, Appl
30	69.5	8.0	368	10	US-09-801-368-308	Sequence 308, App
31	69.5	8.8	805	9	US-10-108-605-113	Sequence 113, App
32	69	8.7	201	10	US-09-901-540-3	Sequence 3, Appl
33	69	8.7	201	10	US-09-896-856-8	Sequence 8, Appl
34	69	8.7	201	10	US-09-901-257-3	Sequence 3, Appl
35	68	8.6	523	10	US-09-799-777-76	Sequence 76, Appl
36	68	8.6	603	10	US-09-906-779-4	Sequence 4, Appl
37	68	8.6	1274	9	US-10-020-215-2	Sequence 2, Appl
38	67.5	8.5	483	9	US-09-738-626-5097	Sequence 5097, Ap
39	67.5	8.5	1170	10	US-09-822-268A-4	Sequence 32, Appl
40	67.5	8.5	1237	9	US-10-024-623-32	Sequence 3998, Ap
41	67	8.4	267	9	US-09-738-626-3998	Sequence 5377, Ap
42	66.5	8.4	342	10	US-09-815-242-5377	Sequence 12283, A
43	66.5	8.4	357	10	US-09-815-242-12283	Sequence 151, App
44	66.5	8.4	626	9	US-09-712-363-151	Sequence 915, App
45	66	8.3	381	9	US-09-764-868-915	

ALIGNMENTS

RESULT 1
US-09-817-514A-2
; Sequence 2, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: french-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2

Alignment Scores:
Pred. No.: 2.39e+45 Length: 2516
Score: 434.50 Matches: 87
Percent Similarity: 72.30% Conservative: 20
Best Local Similarity: 58.78% Mismatches: 38

```
Query Match: 54.79% Indels: 3
DB: 10 Gaps: 1
US-09-856-221-1 (1-449) x US-09-817-514A-2 (1-2516)
QY 13 CACCTTATTCGGCGGATACCTTATTTTCATTCGATGATAACGATTGCTCAGAACCCCGTTTA 72
DB 1888 HsLeuLeuGlyAspLysProLysLeuProLeuSerThrTrpSerAspProLysLeu 1907
QY 73 GAAGAGCCGCCAGTCAACACCATTCGTGATCATATATCAGCATATAAAATGGCGCAACGCGT 132
DB 1908 AspArgAlaAlaAspIleThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927
QY 133 CAACGCGCGCGCTGCGCGCGGAAA-----CGTACTGCAAAATTCGTTAAACGCGCTTG 183
DB 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947
QY 184 TTCCTTCTCAGATAACAAAACACTGCAAGTTACTGGCAGACGCTTAGCACACGCTTA 243
DB 1948 PheLeuProGlnIleAsnGluValMetMetAsnTrpTrpGlnThrLeuAlaGlnArgVal 1967
QY 244 TATAACTTACGTCAATCTGCAATTCGATGCGTACCGCTTGTCAATACCATCTATGCG 303
DB 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987
QY 304 ACACACGACAGTCCGCTGCTACTGCTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
DB 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007
QY 364 GATTGCTCGGACAGTAATGCGGATGACGATGACGATGACGATGACGATGACGATGACGATG 423
DB 2008 LysLeuProGlnSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2027
QY 424 TGGGAGTGGACCACTGATACAG 447
DB 2028 GlyMetValSerGlnLeuThrGln 2035
RESULT 2
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: french-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Alignment Scores:
Pred. No.: 6,32e-43 Length: 2504
Score: 415.00 Matches: 90
Percent Similarity: 71.33% Conservative: 17
Best Local Similarity: 60.00% Mismatches: 41
Query Match: 52.33% Indels: 3
DB: 10 Gaps: 1
US-09-856-221-1 (1-449) x US-09-817-514A-8 (1-2504)
QY 5 CAGGCATCACCATTATTCGGCGGATAA-CCTTATTTTCATTCGATGATAACGATTGTCAGAA 63
DB 1880 GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrTrpAlaAsn 1899
```

```
QY 64 CCGCGTTTAAAGAGCCGCCAGTCAAAACCATTCGTGATCATATATCAGCATATAAAATCGCG 123
DB 1900 ProThrLeuGlyAsnAlaAlaSerLysThrGlnGlnValArgGlnGlnValLeuThr 1919
QY 124 CAACCTCGGT-----CAACGCGCGCGCTTCCGCGGGAACAGTACTGCAAAATTCGTTAAC 177
DB 1920 GlnLeuArgLeuAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1939
QY 178 GCTTTGTTCTCTCAGATATAACAAAAAACTGCAAAAGTTACTGGCAGACGTTAGCACAA 237
DB 1940 AlaLeuPheLeuProGlnGlnAsnSerLysLeuLysGlyTyrTrpArgThrLeuAlaGln 1959
QY 238 CGCCTATAAATTCAGTCATAATCTGACAAATTCATGATGGTCAGCGCTGTTCATTACCATC 297
DB 1960 ArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuSerLeuProLeu 1979
QY 298 TATCGACACACGAGATCCGCTCGTACTGCTTAGTGTGCTGCGTCACCGCTCACAGGC 357
DB 1980 TyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly 1999
QY 358 GGAGGGGATTCCCTCGGACAGTAATCCGATGATCCGATGATCCGATGATGATGATGATGATG 417
DB 2000 GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGluGly 2019
QY 418 GCCAAGTGGGGAGTGCACCAACTGATACAG 447
DB 2020 AlaArgGlyLeuValAsnGlnLeuIleGln 2029
RESULT 3
US-09-808-571A-2
; Sequence 2, Application US/09808571A
; Patent No. US20020106723A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; TITLE OF INVENTION: Receptor for latrotoxin from insects
; FILE REFERENCE: Le A 34 402
; CURRENT APPLICATION NUMBER: US/09/808,571A
; CURRENT FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: DE 100 13 580.3
; PRIOR FILING DATE: 2000-03-18
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1447
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-808-571A-2
Alignment Scores:
Pred. No.: 0.689 Length: 1447
Score: 76.00 Matches: 31
Percent Similarity: 30.71% Conservative: 12
Best Local Similarity: 22.14% Mismatches: 47
Query Match: 9.58% Indels: 50
DB: 10 Gaps: 3
US-09-856-221-1 (1-449) x US-09-808-571A-2 (1-1447)
QY 34 TATTTTTCATTCGATAACGATTCGTGAGAACCCCGTTTAGAAGAAGCGCGCATCAAGCC 93
DB 1196 TyrTyrSer-----TrpLysGlnProSerThrGlyThrGlyLeuLysThr 1211
QY 94 ATTGCTGATCATTTAT----- 108
DB 1212 ProArgGluTyrTyrAsnAsnAlaGlyAlaAlaAlaSerSerProThrGlyAlaArgGly 1231
QY 109 -----CAGCATATAAATCGGCAACTGCGTCAA 135
DB 1232 IleLeuLeuAspSerLysAlaGluGlnArgProGlnTrpGlnLysGluGlyArgArg 1251
QY 136 CGCGCGCGCTTCCGCGGCAACGCTACTGCAAAATTCGTTAACGCTTGTTCCTTCCTCAG 195
DB 1366 ----- 195
```

Db 1252 ArgSerSerArgLeuAlaTyrArgThrAlaAlaAlaSerGlnValLeuPheTyrProSer 1271
QY 196 ATAAACAAAACAACTGCAAGTACTGCAGACGTTAGCACAAACGCTATATACCTACGT 255
Db 1272 TyrLysLys----- 1274
QY 256 CATAACTGACAAATGATGGTCAGCGCTGTGCATTACCCATCTATGCGACACAGAGAT 315
Db 1275 -----ThrLysProGlnProThrGlyTyrProGlnTyrAlaGluAlaLeuAsp 1291
QY 316 CCGTCCGTACTGTTAGTGTGCTGCCTCAGCGCTCAGAGCGGAGGGATTTCGCTCGG 375
Db 1292 ProProLeuAlaThrGlyAsnAlaAlaAlaTyrTyrGlnGlnGlnGlnLeuArgArg 1311
RESULT 4
US-10-151-736-4
; Sequence 4, Application US/10151736
; Publication No. US20020192160A1
; GENERAL INFORMATION:
; APPLICANT: Callaghan, Michelle J.
; APPLICANT: Sutherland, Lindfield
; TITLE OF INVENTION: No. US20020192160A1 Human Tumour Suppressor Gene
; FILE REFERENCE: RICE-01000N
; CURRENT APPLICATION NUMBER: US/10/151,736
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 09/403,402
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: PCT/AU98/00280
; PRIOR FILING DATE: 1998-04-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2799
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-151-736-4
Alignment Scores:
Pred. No.: 0.898 Length: 2799
Score: 76.00 Matches: 33
Percent Similarity: 43.59% Conservative: 18
Best Local Similarity: 28.21% Mismatches: 46
Query Match: 9.58% Indels: 20
Gaps: 5
US-09-856-221-1 (1-449) x US-10-151-736-4 (1-2799)
QY 43 TTGGATAACGATTGTGTACAGACCCGTTTAGAAGAGCGCGCACTCAACACCTTCGTGAT 102
Db 1409 LeuGlnAsnLysTyrThrProGlyArgArgGluGluAlaAlaValThr----- 1425
QY 103 CATATCAGCATATAATGCGCAACTGCGTCAACGCGG-----GCCTTGGCG 150
Db 1426 -----MetArgPheLeuArgSerValAlaArgValPheValIleLeuSer 1440
QY 151 GCGNAACGCTACTGCAAAATTCCTTAACCGCTTGTTCCTCTCAGATAAACAACAACTG 210
Db 1441 ValGluMetAlaSerSerLysLysAsnAsnPheIleProGlnProLysGlyLysCys 1460
QY 211 CAAAGTTACTGCGACAGCTTA-----GCACACACGCTATATAACTTTACGTCAT 258
Db 1461 LysArgValPheGlnAlaLeuLeuProTyrAlaValGluGluCysAsnValAlaGlu 1480
QY 259 AATCTGCAATGATGGTCAGCGCTTGTCAATACCATCTATGCGACACAGCAGATCGG 318
Db 1481 SerLeuIleVal-----ProValArgMetGlyIle---AlaArgProThrAlaPro 1496
QY 319 TCGGTACTGTTAGTGTGCTGCCTCAGCGCTCAGAGCGGAGGGATTG 369
Db 1497 PheThrLeuAlaSerThrSerIleAspAlaMetGlnGlySerGluGluLeu 1513
RESULT 5

US-09-933-638A-12
; Sequence 12, Application US/09933638A
; Patent No. US20020160952A1
; GENERAL INFORMATION:
; APPLICANT: Kazantsev, Aleksey G.
; APPLICANT: Thompson, Leslie M.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: INHIBITION OF PROTEIN-PROTEIN INTERACTION
; FILE REFERENCE: 01997-289001
; CURRENT APPLICATION NUMBER: US/09/933,638A
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 60/226,502
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-933-638A-12
Alignment Scores:
Pred. No.: 0.783 Length: 338
Score: 73.50 Matches: 23
Percent Similarity: 42.72% Conservative: 21
Best Local Similarity: 22.33% Mismatches: 46
Query Match: 9.27% Indels: 13
Gaps: 2
US-09-856-221-1 (1-449) x US-09-933-638A-12 (1-338)
QY 73 GAAGAAGCCGCCAGTCAACACCTTCGTGATCATATACGATCAAAATCGGCACTGCGT 132
Db 73 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 92
QY 133 CAACGCGCGCTTCGCGCGAAGCTGCTGCAAAATTCGTTAAACGCTTTGTTCCCTCT 192
Db 93 GlnGlnGlnAlaValAlaAlaAlaAlaValGlnGlnSerThrSer----- 107
QY 193 CAGATAACAAAAAACTGCAAAAGTTACTGGCAGAGCTTAGCACACGCTATATAACTTA 252
Db 108 -----GlnGlnAlaThrGlnGlyThrSerGlyGlnAlaProGlnLeu 121
QY 253 CGTCATAATCTCACAATGTGATGTCACCGCTTG-----TCATTACCATCTATGCGACA 306
Db 122 PheHisSerGlnThrLeuThrAlaProLeuProGlyThrThrProLeuTyrProSer 141
QY 307 CCAGCAGATCCGTCGCTACTGCTTAGTGTGCTGCGCTCACAGCGCGGAGGGAT 366
Db 142 ProMetThrProMetThrProIleThrProAlaThrProAlaSerGluSerGlyIle 161
QY 367 TTGCTCTCGG 375
Db 162 ValProGln 164
RESULT 6
US-09-849-243-16
; Sequence 16, Application US/09849243
; Patent No. US20020157127A1
; GENERAL INFORMATION:
; APPLICANT: Kirschbaum, Bernd
; APPLICANT: Berglund, Erick
; APPLICANT: Meisterer, Michael
; APPLICANT: Polites, Greg
; TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
; COMPLEXES FROM TRANSGENIC
; NON-HUMAN ANIMALS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HELLER, EHRMAN, WHITE & MCAULIFFE
; STREET: 1666 K Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.

COUNTRY: USA
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

US-09-856-221-1 (1-449) x US-09-849-243-16 (1-371)

[illegible]

RESULT 7

```

US-09-944-413-55
; Sequence 55, Application US/09944413
; Patent NO. US20020156004A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerlits, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; PRIOR FILING DATE: December 16, 1999
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020156004A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020156004A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999

```

```
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-413-55

Alignment Scores:
Pred. No.: 1.86 Length: 694
Score: 71.50 Matches: 27
Percent Similarity: 45.54% Conservatives: 24
Best Local Similarity: 24.11% Mismatches: 38
Query Match: 9.02% Indels: 23
DB: Gaps: 4

US-09-856-221-1 (1-449) x US-09-944-413-55 (1-694)
QY 58 TCAGAGCCCGTTTGAAGAGCGCCGAGTCACCAACCTTCGTGATCATTTATCAGCATAA 117
DB 118 SerGlyAspArgValGlyGluLysArgAsnLysThrThrGluGluAsnGlyGluLysGly 137
QY 118 ATCGGGCAACTCGTCACAGCGCGCGCTTCGCGGCGAAGACTGCTGCAAAATCGTTAAAC 177
DB 138 ThrGluIlePheArgAlaSerAlaValIleProSerLysAspLysAla----- 153
QY 178 GCTTGTGCTCTCCCTCAGATAAACAAAAAAGTCAAAAGTTACTGGCAGAGCTTAGCACAA 237
DB 154 AlaPhePheLeu-----SerTyrGluGluLeuLeuGlnArg 165
QY 238 CGCCTATATAACTAGTCATATCTGACAAATGATGTCAGCGCTGTCATTACCCATC 297
DB 166 ArgLeuGlyLysTyrGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185
QY 298 TATCGCACACAGCAGATCGCTCGTACTGCTAGTGTGCGCGTACCCGCTCACAGGC 357
DB 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201
QY 358 GGAGGGGATTGCTCGCAGACAGTAATCCCGATGATC 393
DB 202 -----ValLeuProLeuHis 206

RESULT 8
US-09-944-403-55
; Sequence 55, Application US/09944403
; Patent No. US20020165143A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fillaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
```

; PRIOR FILING DATE: February 22, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/05841
 ; PRIOR FILING DATE: March 2, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/08439
 ; PRIOR FILING DATE: March 30, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/14042
 ; PRIOR FILING DATE: May 22, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/20710
 ; PRIOR FILING DATE: July 28, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/32678
 ; PRIOR FILING DATE: December 1, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US01/06520
 ; PRIOR FILING DATE: February 28, 2001
 ; NUMBER OF SEQ ID NOS: 120
 ; SEQ ID NO 55
 ; LENGTH: 694
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-09-944-403-55

Alignment Scores:
 Pred. No.: 1.86 Length: 694
 Score: 71.50 Matches: 27
 Percent Similarity: 45.54% Conservative: 24
 Best local Similarity: 24.11% Mismatches: 38
 Query Match: 9.02% Indels: 23
 Gaps: 4

US-09-856-221-1 (1-449) x US-09-944-403-55 (1-694)

QY 58 TCAGAACCCCTTGTAGAGAGCGCGCAGTCAACATTCGTGATCATTCATTCAGCATAAA 117
 Db 118 SerGlyAspArgValLysGluLysAAsnLysThrThrGluGluAsnGlyGluLysGly 137
 QY 118 ATCGGCAACTCGTCAACGCGCGCTCCGCGGCAACGCTACTGCAAAATTCGTTAAC 177
 Db 138 ThrGluIlePheArgAlaSerAlaValIleProSerLysAspLysAla----- 153
 QY 178 GCTTTCTCTCTCTCCTCAGATAAACAACAACTGCAAAAGTTACTGCGCAGAGCTTAGCACA 237
 Db 154 AlaPhePheLeu-----SerTyrGluGluLeuLeuGlnArg 165
 QY 238 CGCCTATATACCTTACGTATATCTACAAATTCACAAATTCGTCAGCGCTGTCTATACCCATC 297
 Db 166 ArgLeuGlyLysTyrGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185
 QY 298 TATGCGACACACGACGATCGCTCGTACTGCTTGTAGTGTGCGTCACGCGCTCACAGGC 357
 Db 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201
 QY 358 GGAGGGGATTGCTCGGACAGATGATCGCATGTAC 393
 Db 202 -----ValLeuProLeuHis 206

RESULT 9

US-09-944-896-55

; Sequence 55, Application US/09944896

; Patent No. US20020168715A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin

; APPLICANT: Batson, David

; APPLICANT: Batson, Dan

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gerriksen, Mary

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul

; APPLICANT: Grimaldi, Christopher

; APPLICANT: Gurney, Austin

; APPLICANT: Hillen, Kenneth

; APPLICANT: Kljavin, Ivar

; APPLICANT: Napier, Mary

; APPLICANT: Roy, Margaret

; APPLICANT: Tumas, Daniel
 ; APPLICANT: Wood, William
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P2548F1C1
 ; CURRENT FILING DATE: 2001-08-31
 ; PRIOR APPLICATION NUMBER: 09/866,028
 ; PRIOR FILING DATE: 2001-05-25
 ; PRIOR APPLICATION NUMBER: 60/069,334
 ; PRIOR FILING DATE: December 11, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,335
 ; PRIOR FILING DATE: December 11, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,278
 ; PRIOR FILING DATE: December 11, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,425
 ; PRIOR FILING DATE: December 12, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,696
 ; PRIOR FILING DATE: December 16, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,694
 ; PRIOR FILING DATE: December 16, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,702
 ; PRIOR FILING DATE: December 16, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,870
 ; PRIOR FILING DATE: December 17, 1997
 ; PRIOR APPLICATION NUMBER: 60/069,873
 ; PRIOR FILING DATE: December 17, 1997
 ; PRIOR APPLICATION NUMBER: 60/068,017
 ; PRIOR FILING DATE: December 18, 1997
 ; PRIOR APPLICATION NUMBER: 60/070,440
 ; PRIOR FILING DATE: January 5, 1998
 ; PRIOR APPLICATION NUMBER: 60/074,086
 ; PRIOR FILING DATE: February 9, 1998
 ; PRIOR APPLICATION NUMBER: 60/074,092
 ; PRIOR FILING DATE: February 9, 1998
 ; PRIOR APPLICATION NUMBER: 60/075,945
 ; PRIOR FILING DATE: February 25, 1998
 ; PRIOR APPLICATION NUMBER: 60/112,850
 ; PRIOR FILING DATE: December 16, 1998
 ; PRIOR APPLICATION NUMBER: 60/113,296
 ; PRIOR FILING DATE: December 22, 1998
 ; PRIOR APPLICATION NUMBER: 60/146,222
 ; PRIOR FILING DATE: July 28, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US98/19330
 ; PRIOR FILING DATE: September 16, 1998
 ; PRIOR APPLICATION NUMBER: PCT/US98/25108
 ; PRIOR FILING DATE: December 1, 1998
 ; PRIOR APPLICATION NUMBER: 09/216,021
 ; PRIOR FILING DATE: December 16, 1998
 ; PRIOR APPLICATION NUMBER: 09/218,517
 ; PRIOR FILING DATE: December 22, 1998
 ; PRIOR APPLICATION NUMBER: 09/254,311
 ; PRIOR FILING DATE: March 3, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US99/12252
 ; PRIOR FILING DATE: June 22, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: September 15, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US99/28409
 ; PRIOR FILING DATE: NO. US20020168715A1ember 30, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: NO. US20020168715A1ember 30, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US99/28301
 ; PRIOR FILING DATE: December 1, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: December 16, 1999
 ; PRIOR APPLICATION NUMBER: PCT/US00/03565
 ; PRIOR FILING DATE: February 11, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414
 ; PRIOR FILING DATE: February 22, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/05841
 ; PRIOR FILING DATE: March 2, 2000
 ; PRIOR APPLICATION NUMBER: PCT/US00/08439
 ; PRIOR FILING DATE: March 30, 2000

```
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 55
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-896-55

Alignment Scores:
Pred. No.: 1.86 Length: 694
Score: 71.50 Matches: 27
Percent Similarity: 45.54% Conservative: 24
Best Local Similarity: 24.11% Mismatches: 38
Query Match: 9.02% Indels: 23
Dbs: 9 Gaps: 4

US-09-856-221-1 (1-449) x US-09-944-896-55 (1-694)
QY 58 TCAGAACCCGGTTAGAGAGCGCCAGTCAACCATTCGTGATCATTTATCAGCATATAA 117
Db 118 SerGlyAspArgValGlyGluLysArgAsnLysThrThrGluGluAsnGlyGluLysGly 137
QY 118 ATCGGCAACTCGTCAACGCGCGCTTGGCGGGAACGACTGCAAAATTCGTTAACCC 177
Db 138 ThrGluLeuPheArgAlaSerAlaValIleProSerLysAspLysAla----- 153
QY 178 GCATTGTTCCCTCCACATATAACAAAACTGCAAAAGTTACTGGCAGACGTTAGCACAA 237
Db 154 AlaPhePheLeu-----SerTyrGluGluLeuLeuGluGlnArg 165
QY 238 CGCCTATATAACTTACGTCAATATCTGACAAATGATGTCGACCGGTGTGTCATTACCCATC 297
Db 166 ArgLeuGlyLysTyrGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185
QY 298 TATGCGACACGACGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
Db 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201
QY 358 GGAGGGGATTCCTCGGACAGTAAATGCGGATGTC 393
Db 202 -----ValLeuProLeuHis 206

RESULT 10
US-09-944-944-55
; Sequence 55, Application US/09944944
; Patent No. US20020173463A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerlitsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavio, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2548P1C1

; CURRENT APPLICATION NUMBER: US/09/944, 944
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866, 028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067, 411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069, 334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069, 335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069, 278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069, 425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069, 696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069, 694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069, 702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069, 870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069, 873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068, 017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070, 440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074, 086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074, 092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075, 945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112, 850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113, 296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146, 222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216, 021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218, 517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254, 311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020173463A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020173463A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
```


Search completed: January 15, 2003, 15:58:34
Job time : 13.4233 secs

[illegible]

Alignment Scores:		
Pred. No.:	1.86	Length:
Score:	71.50	Matches:
Percent Similarity:	45.54%	Conservative:
Best Local Similarity:	24.11%	Mismatches:
Query Watch:	9.02%	Indels:
Ds:	10	Gaps:
	4	

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:44:53 ; Search time 20.6921 seconds
(without alignments)
4172.064 Million cell updates/sec

Title: US-09-856-221-1
Perfect score: 793
Sequence: 1 tgtcagcactcacttat.....gtgacccaactgatacagtt 449

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=rlp
-O=/cgn2_1/USFTQ_SPOOL/US09856221/runat_15012003_153925_12500/app_query.fasta_1.1.1948
-DB=PIR_73 -OFMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -IOOPEXT=0
-UNIT5=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMI=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221@cgn_1_1_95@runat_15012003_153925_12500 -NCPU=3
-NO_XLPRY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=7 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84.5	10.7	270	2 AB1858	tryptophan synthas
2	84.5	10.7	568	2 C72129	probable outer mem
3	84.5	10.7	568	2 G86493	probable leader pe
C 4	84	10.8	317	2 AD3067	hypothetical prote
C 5	84	10.8	338	2 F98219	hypothetical prote
C 6	78	9.8	640	2 T41977	hypothetical prote
C 7	77.5	9.9	307	2 B69837	hypothetical prote
8	76	9.6	2555	2 C69681	peptide synthetase
9	75.5	9.5	622	2 T10009	probable serine/th
10	75	9.5	316	1 S34437	transcription init
C 11	75	9.6	635	2 F96660	protein F2K11.10
12	74.5	9.4	407	2 A95155	hypothetical prote
13	74.5	9.4	1437	2 T31093	probable protein-t
14	74	9.3	1186	2 T19050	hypothetical prote

15	73.5	9.3	297	1	I51648	transcription init
16	73.5	9.3	339	1	TMH2D	transcription init
17	73.5	9.3	450	2	AG0222	flagellar hook-len
18	73.5	9.3	622	2	H86910	probable serine/th
19	73.5	9.3	1718	1	JQ1734	genome polyprotein
20	73	9.2	277	2	S56591	hypothetical 30.5K
21	72.5	9.1	936	2	D97630	hypothetical prote
22	72.5	9.1	1078	2	PC4198	peptide synthetase
23	72.5	9.1	2688	2	I49477	alpha-A-crystallin
24	72	9.1	442	2	A41015	dihydrolipoamide S
25	72	9.1	488	2	T38481	probable pre-mRNA
26	72	9.1	881	2	E82097	protein-P-II uridy
27	72	9.1	1145	2	S13643	PRP22 protein-ye
28	71.5	9.0	201	2	A30833	ribosomal protein
29	71.5	9.0	268	2	C83104	hypothetical prote
30	71.5	9.0	470	1	A25685	hypothetical prote
31	71.5	9.0	470	2	T39796	negative sexual co
32	71.5	9.0	820	2	T46412	ubiquitin-protein
33	71	9.0	275	2	AC2293	hypothetical prote
34	71	9.0	302	1	JC4059	transcription init
35	71	9.0	317	2	T39736	hypothetical prote
36	71	9.0	380	2	S75348	hypothetical prote
37	71	9.0	509	2	T16846	hypothetical prote
38	71	9.0	886	2	A85905	hypothetical prote
39	71	9.0	886	2	G85036	hypothetical prote
40	71	9.0	2441	2	S39161	CREB-binding prote
41	70.5	8.9	365	2	T24955	hypothetical prote
42	70.5	8.9	694	2	A69768	transcription anti
43	70.5	8.9	901	2	F83781	transposase (08) /
44	70	8.8	303	2	A11070	probable membrane
45	70	8.8	321	2	T47280	transposase-like p

ALIGNMENTS

RESULT 1

AB1858
tryptophan synthase alpha chain [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AB1858
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriku
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB1858
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <KUR>
A:Cross-references: GB:HA000019; PIDN:BA072369.1; PID:gl7129756; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: trpA
C:Superfamily: tryptophan synthase alpha chain; tryptophan synthase alpha chain homol

Alignment Scores:	Pred. No.:	Length:
Score:	1.19	270
Percent Similarity:	84.50	Matches: 40
Best Local Similarity:	46.96%	Conservative: 14
Query Match:	34.78%	Mismatches: 34
DB:	10.66%	Indels: 27
	2	Gaps: 8
US-09-856-221-1 (1-449) x AB1858 (1-270)		
Qy 91	ACCATCTGTGATCATATTCAGCATATAATCGCGCACTGCGTCAACGCGCGCTTGGCG	150
Db 3	ThrlleSerAspHisPheGln---SerLeuArgGlnArgGlnCysAlaLeuilePro 21	
Qy 151	GCGAACGCTACTGCAAAATTCGTTAACCGCTTTTCTCTCAGATAAACAACAACTG 210	

```
Db 22 -----PheIleThrAlaGlyAspProAspLeu 30
QY 211 CAAAGTTACTGCGAGCGTTAGCACACGCTATATAACTTA---CGTCATAAFTCTGACA 267
Db 31 -----GlnThrThrAlaGluAlaLeuArgileLeuAspArgHisGly----- 44
QY 268 ATTGATGGTCAGCGGTTGTCATTACCGCATATCGCACACCA---GCAGATCCGTCGGTA 324
Db 45 AlaAspPheIleGluLeuGlyValPro---TyrSerAspProLeuAlaAspGlyProVal 63
QY 325 CTGCTTAGTCTGCCGTCACCGCTCACAGGCGGAGGGGATTTCGCCCTCGG----- 375
Db 64 IleGlnAlaAlaAlaThrArgAlaLeuGlnArgGlyThrLysLeuGluGlnValLeuGlu 83
QY 376 -----ACAGTAATGCCGATGACCGTTTCCTCCGATTATTCGTG 411
Db 84 MetLeuGluThrValIleProSerLeuGlnAlaProIleLeu 98

RESULT 2
C72129
probable outer membrane leader peptide (omp) CPn0021 - Chlamydothilla pneumoniae (strains
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: C72129; B81542
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72129
A:Molecule type: DNA
A:Residues: 1-568 <ARN>
A:Cross-references: GB:AE001587; GB:AE001363; NID:g4376271; PIDN:AAID18174.1; PID:g437627
A:Experimental source: strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: B81542
A:Molecule type: DNA
A:Residues: 1-568 <REA>
A:Cross-references: GB:AE002234; GB:AE002161; NID:g7189667; PIDN:AAF38558.1; PID:g718967
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CPn0021; CP0755

Alignment Scores:
Pred. No.: 1.32 Length: 568
Score: 84.50 Matches: 40
Percent Similarity: 42.66% Conservative: 21
Best local Similarity: 27.97% Mismatches: 55
Query Match: 10.66% Indels: 27
DB: 2 Gaps: 6

US-09-856-221-1 (1-449) x C72129 (1-568)
QY 58 TCAGAACCCCGTTTAGAACAGCC-----GCCAGTCAAAACCATTCGTGAT----- 102
Db 232 AlaSerProGlnAspGlnGluAlaLeuTyAlaLeuGlyLysLeuLysAspGlyGln 251
QY 103 CATTATCAGCATAAATGGCGCACTGGTCACAGCGCGGCTTCCCGCGGAACGTACT 162
Db 252 SerTyrTyrAsnIleLysGlnLeuGlnLysProAspValThrLeuAlaAla 271
QY 163 GCAAAATTCGTTAACCGCTTTG-----TTCCCTTCCTCAGATAACAAAAA 207
Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluAlaAspAlaLeuProValIleLysLys--- 290
QY 208 CTGCAAAAGTTACTGGCAGACGTTAGCACACGCC-----CTATATAACTTACGTGAT 258
Db 291 -----GlnAlaLeuGluArgProArgAlaLeuTyAlaLeuArgHis 305
QY 103 CATTATCAGCATAAATGGCGCACTGGTCACAGCGCGGCTTCCCGCGGAACGTACT 162
Db 252 SerTyrTyrAsnIleLysGlnLeuGlnLysProAspValThrLeuAlaAla 271
QY 163 GCAAAATTCGTTAACCGCTTTG-----TTCCCTTCCTCAGATAACAAAAA 207
Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluAlaAspAlaLeuProValIleLysLys--- 290
QY 208 CTGCAAAAGTTACTGGCAGACGTTAGCACACGCC-----CTATATAACTTACGTGAT 258
Db 291 -----GlnAlaLeuGluArgProArgAlaLeuTyAlaLeuArgHis 305
QY 259 AATCTGACAATGATGGTCAGCGGTTGTGTCATTATCCCATCTATCGCACACCATGATCCG 318
Db 305 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 319 TCCGTACTGCTTAGTCTGCTCCGCTCACAGCGGCGGAGGGGATTTCGCCCT- 372
Db 325 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 319 TCCGTACTGCTTAGTCTGCTCCGCTCACAGCGGCGGAGGGGATTTCGCCCT- 372
Db 325 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 373 -----CGGACAGTAATGCCGATGACCGATTTCCTCGGAA 414
Db 346 LeuGluTyrIleThrGluArgLeuValGlnProHisTyrAsnGluThrLeuAlaLeuSer 365
QY 415 AATGCCAAG 423

RESULT 3
G86493
probable leader peptide Omp [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G86493
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: G86493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <STO>
A:Cross-references: GB:BA000008; NID:g8978396; PIDN:BA098233.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPj0021

Alignment Scores:
Pred. No.: 1.32 Length: 568
Score: 84.50 Matches: 40
Percent Similarity: 42.66% Conservative: 21
Best local Similarity: 27.97% Mismatches: 55
Query Match: 10.66% Indels: 27
DB: 2 Gaps: 6

US-09-856-221-1 (1-449) x G86493 (1-568)
QY 58 TCAGAACCCCGTTTAGAACAGCC-----GCCAGTCAAAACCATTCGTGAT----- 102
Db 232 AlaSerProGlnAspGlnGluAlaLeuTyAlaLeuGlyLysLeuLysAspGlyGln 251
QY 103 CATTATCAGCATAAATGGCGCACTGGTCACAGCGCGGCTTCCCGCGGAACGTACT 162
Db 252 SerTyrTyrAsnIleLysGlnLeuGlnLysProAspValThrLeuAlaAla 271
QY 163 GCAAAATTCGTTAACCGCTTTG-----TTCCCTTCCTCAGATAACAAAAA 207
Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluAlaAspAlaLeuProValIleLysLys--- 290
QY 208 CTGCAAAAGTTACTGGCAGACGTTAGCACACGCC-----CTATATAACTTACGTGAT 258
Db 291 -----GlnAlaLeuGluArgProArgAlaLeuTyAlaLeuArgHis 305
QY 259 AATCTGACAATGATGGTCAGCGGTTGTGTCATTATCCCATCTATCGCACACCATGATCCG 318
Db 305 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 319 TCCGTACTGCTTAGTCTGCTCCGCTCACAGCGGCGGAGGGGATTTCGCCCT- 372
Db 325 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 319 TCCGTACTGCTTAGTCTGCTCCGCTCACAGCGGCGGAGGGGATTTCGCCCT- 372
Db 325 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 373 -----CGGACAGTAATGCCGATGACCGATTTCCTCGGAA 414
Db 346 LeuGluTyrIleThrGluArgLeuValGlnProHisTyrAsnGluThrLeuAlaLeuSer 365
QY 415 AATGCCAAG 423
```

```
Db 22 -----PheIleThrAlaGlyAspProAspLeu 30
QY 211 CAAAGTTACTGCGAGCGTTAGCACACGCTATATAACTTA---CGTCATAAFTCTGACA 267
Db 31 -----GlnThrThrAlaGluAlaLeuArgileLeuAspArgHisGly----- 44
QY 268 ATTGATGGTCAGCGGTTGTCATTACCGCATATCGCACACCA---GCAGATCCGTCGGTA 324
Db 45 AlaAspPheIleGluLeuGlyValPro---TyrSerAspProLeuAlaAspGlyProVal 63
QY 325 CTGCTTAGTCTGCCGTCACCGCTCACAGGCGGAGGGGATTTCGCCCTCGG----- 375
Db 64 IleGlnAlaAlaAlaThrArgAlaLeuGlnArgGlyThrLysLeuGluGlnValLeuGlu 83
QY 376 -----ACAGTAATGCCGATGACCGTTTCCTCCGATTATTCGTG 411
Db 84 MetLeuGluThrValIleProSerLeuGlnAlaProIleLeu 98

RESULT 2
C72129
probable outer membrane leader peptide (omp) CPn0021 - Chlamydothilla pneumoniae (strains
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: C72129; B81542
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72129
A:Molecule type: DNA
A:Residues: 1-568 <ARN>
A:Cross-references: GB:AE001587; GB:AE001363; NID:g4376271; PIDN:AAID18174.1; PID:g437627
A:Experimental source: strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: B81542
A:Molecule type: DNA
A:Residues: 1-568 <REA>
A:Cross-references: GB:AE002234; GB:AE002161; NID:g7189667; PIDN:AAF38558.1; PID:g718967
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CPn0021; CP0755

Alignment Scores:
Pred. No.: 1.32 Length: 568
Score: 84.50 Matches: 40
Percent Similarity: 42.66% Conservative: 21
Best local Similarity: 27.97% Mismatches: 55
Query Match: 10.66% Indels: 27
DB: 2 Gaps: 6

US-09-856-221-1 (1-449) x C72129 (1-568)
QY 58 TCAGAACCCCGTTTAGAACAGCC-----GCCAGTCAAAACCATTCGTGAT----- 102
Db 232 AlaSerProGlnAspGlnGluAlaLeuTyAlaLeuGlyLysLeuLysAspGlyGln 251
QY 103 CATTATCAGCATAAATGGCGCACTGGTCACAGCGCGGCTTCCCGCGGAACGTACT 162
Db 252 SerTyrTyrAsnIleLysGlnLeuGlnLysProAspValThrLeuAlaAla 271
QY 163 GCAAAATTCGTTAACCGCTTTG-----TTCCCTTCCTCAGATAACAAAAA 207
Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluAlaAspAlaLeuProValIleLysLys--- 290
QY 208 CTGCAAAAGTTACTGGCAGACGTTAGCACACGCC-----CTATATAACTTACGTGAT 258
Db 291 -----GlnAlaLeuGluArgProArgAlaLeuTyAlaLeuArgHis 305
QY 259 AATCTGACAATGATGGTCAGCGGTTGTGTCATTATCCCATCTATCGCACACCATGATCCG 318
Db 305 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 319 TCCGTACTGCTTAGTCTGCTCCGCTCACAGCGGCGGAGGGGATTTCGCCCT- 372
Db 325 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 319 TCCGTACTGCTTAGTCTGCTCCGCTCACAGCGGCGGAGGGGATTTCGCCCT- 372
Db 325 LeuProSerGluIleGlyLeuProIleAlaLeuProIlePheLeuLysThrLysAsnSer 325
QY 373 -----CGGACAGTAATGCCGATGACCGATTTCCTCGGAA 414
Db 346 LeuGluTyrIleThrGluArgLeuValGlnProHisTyrAsnGluThrLeuAlaLeuSer 365
QY 415 AATGCCAAG 423
```

```

Db 366 PheSerLys 368
:::|
RESULT 4
AD3067
hypothetical protein Atu4154 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD3067
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
S.E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD3067
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-317 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA44954.1; PID:g17742609; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4154
A:Map position: linear chromosome

Alignment Scores:
Pred. No.: 1.38 Length: 317
Score: 84.00 Matches: 42
Percent Similarity: 43.70% Conservative: 17
Best Local Similarity: 31.11% Mismatches: 53
Query Match: 10.76% Indels: 23
DB: 2 Gaps: 8

US-09-856-221-1 (1-449) x AD3067 (1-317)
QY 443 ATCAGTTGGTCACT-----CCCCACTTGGCATTTTCCAGATAATCGGAAAA 396
Db 62 ValSerTrpValGluPheThrAlaGlyProLeuValGluAlaLeuAsnValGlySer 81
QY 395 CGGTACATCGGCATTACTGTCGAGCAAAATCCCTCCG----- 357
Db 82 IleAsnValGlyTrpThr-----GlyAspAlaProIlePheGlyGlnAlaAlaGly 99
QY 356 CCTGTGAGCGGTGACGGCGACTAAGCAGTACGACGATCTGCTGTGTCGCATAG 297
Db 100 SerAlaIleValTyValAlaAlaLeuProSer---AsnGlyLysGlyGluAlaIlePhe 118
QY 296 ATGGGTAAATGACAACGGCTGACCATCAATTGTCAGATTATGACGTAAGTTATATAGCGGT 237
Db 119 ThrLysProGluSerGlyIleLysSerVal-AlaAspLysGlyLysValGlyVa 138
QY 236 TGTGCTA---ACGTCTGCAGTACTTTCGAGTTTTCGTTTCTGAGGAAGCAAAA 180
Db 138 lGlyLysGlyThrSerAlaHisAsnLeuValAlaAlaLeuGluLysAsnGlyLeuLy 158
QY 179 GCGGTTAACGAATTTGCAG-----TACGTTTCGCGGCAAGCGCGCGTTCGACGCACT 126
Db 158 s---LeuSerAspIleAspValThrTyLeuSerProAla-----AspAlaAl 173
QY 125 TGGCGCATTTTATGCTGATATGATCAGGAATGTTTGACTGG 83
Db 173 aAlaAlaPheAlaSerAspLysIleAspAlaTrpAlaValTrp 187

RESULT 5
F98219
hypothetical protein AGR_L1399 [imported] - Agrobacterium tumefaciens (strain C58, Cere)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: F98219
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.

```

```

Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: F98219
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89280.1; PID:g15159111; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1399
A:Map position: linear chromosome

Alignment Scores:
Pred. No.: 1.39 Length: 338
Score: 84.00 Matches: 42
Percent Similarity: 43.70% Conservative: 17
Best Local Similarity: 31.11% Mismatches: 53
Query Match: 10.76% Indels: 23
DB: 2 Gaps: 8

US-09-856-221-1 (1-449) x F98219 (1-338)
QY 443 ATCAGTTGGTCACT-----CCCCACTTGGCATTTTCCAGATAATCGGAAAA 396
Db 83 ValSerTrpValGluPheThrAlaGlyProLeuValGluAlaLeuAsnValGlySer 102
QY 395 CGGTACATCGGCATTACTGTCGAGCAAAATCCCTCCG----- 357
Db 103 IleAsnValGlyTrpThr-----GlyAspAlaProIlePheGlyGlnAlaAlaGly 120
QY 356 CCTGTGAGCGGTGACGGCGACTAAGCAGTACGACGATCTGCTGTGTCGCATAG 297
Db 121 SerAlaIleValTyValAlaAlaLeuProSer---AsnGlyLysGlyGluAlaIlePhe 139
QY 296 ATGGGTAAATGACAACGGCTGACCATCAATTGTCAGATTATGACGTAAGTTATATAGCGGT 237
Db 140 ThrLysProGluSerGlyIleLysSerVal-AlaAspLysGlyLysValGlyVa 159
QY 236 TGTGCTA---ACGTCTGCAGTAACTTTCGAGTTTTCGTTTCTGAGGAAGCAAAA 180
Db 159 lGlyLysGlyThrSerAlaHisAsnLeuValAlaAlaLeuGluLysAsnGlyLeuLy 179
QY 179 GCGGTTAACGAATTTGCAG-----TACGTTTCGCGGCAAGCGCGCGTTCGACGCACT 126
Db 179 s---LeuSerAspIleAspValThrTyLeuSerProAla-----AspAlaAl 194
QY 125 TGGCGCATTTTATGCTGATATGATCAGGAATGTTTGACTGG 83
Db 194 aAlaAlaPheAlaSerAspLysIleAspAlaTrpAlaValTrp 208

RESULT 6
T41977
hypothetical protein U76 - human herpesvirus 7 (strain JI)
C:Species: human herpesvirus 7
A:Variety: strain JI
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T41977
R:Nicholas, J.
submitted to the EMBL Data Library, December 1995
A:Description: Determination and analysis of the complete nucleotide sequence of huma
A:Reference number: 222022
A:Accession: T41977
A:Status: preliminary; translated from CB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-640 <NIC>
A:Cross-references: EMBL:U03400; PIDN:AA054737.1
A:Experimental source: strain JI
C:Genetics:
A:Note: U76
C:Superfamily: varicella-zoster virus gene 54 protein

Alignment Scores:
Pred. No.: 6.59 Length: 640

```

US-09-856-221-1 (1-449) x T41977 (1-640)

8 JUL 1968

peptide synthetase ppsC - Bacillus subt

C;Species: *Bacillus subtilis*

C;Accession: C69681; I40458; S49135

C.; Bron, S.; Brouillet, S.; Brusch, C.

Nature 330, 243-250, 1993.
A: Authors: Foulger, D.; Fritz, C.; Futi

Koetter, P.; Koningstein, G.; Krogh, S.

Y, M.: Ogawa, K.; Ogiwara, A.; Oudega,

A; Authors: Schleich, S.; Schroeter, R.;

T.; WILKES, F.: Wipac, A.; Yamamoto, E.
Authors: Yoshikawa, H.F.: Zumstein, E.

A: Reference number: A69580; MUID: 980440

A; Status: preliminary; nucleic acid seq

A;Residues: 1-2555 <KUN>

A; Experimental source: strain Ib8

Microbiology 141, 643-646, 1955

A: Accession: T40458

A; Molecule type: DNA

A; Cross-references: EMBL:Z34883; NID:95

A;Gene: ppsC; pps3

E: 510-953/Domain: acetate-CoA ligase

F:1551-1990/Domain: acetate-CoA ligase

F;1002,2038/Binding site: phosphopante

Alignment Scores:

Score: 76.00

BEST LOCAL SIMULTANEITY: 24:57.0

A:Residues: 1-407 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75434.1; PID:g14972819; GSPDB:GN00164; TIGR:
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SPI336

Alignment Scores:
 Pred. No.: 14.7 Length: 407
 Score: 74.50 Matches: 35
 Percent Similarity: 38.85% Conservative: 26
 Best Local Similarity: 22.29% Mismatches: 54
 Query Match: 9.39% Indels: 42
 DB: 7 Gaps: 7

US-09-856-221-1 (1-449) x A95155 (1-407)

QY 66 CCGTTTAGAAGAGCGCCAGTCAGTCAACCATTCATCATCATATAAATCGCGCA 125
 Db 179 PropheargArgGluGlyGlnAlaThrAsnSerGluThrLeuLysAlaLeuGlyAsnLeu 198
 QY 126 ACTCGTCAACGCGCGCTTCCGCGGGAACGTAC----- 161
 Db 199 AsnProSerArgSerGlyMetSerGlyLysValTyTySerGluGlyLeuAlaProThr 218
 QY 162 -----TGCAAAATTCGTTAACCGC 179
 Db 219 LeuValArgGlyLysGlyGluGlyPhelLysValAlaIleProCysMetThrProAspArg 238
 QY 180 TTTGTTCTCTCCAGATAAACAACAACTGCAAGTTACTGGCAGACGTTAGCACACG 239
 Db 239 Leu-----AspLysArgGlnAsnGlyArgGlyPhelLysGluAsnGln-GluPr 254
 QY 240 CCTATATACTTA-----CGTCAATACTGACAAATGATGGTCAG---CCGTT 284
 Db 254 CMetPheThrLeuAsnThrGlnAspArgTyGlyIleValValGlyAspLeuProTh 274
 QY 285 GTCATTAACCC-----ATCTATGCGACACACAGCAGATCCGTCGTACTGCTTAG 332
 Db 274 rSerPhelLysGluThrGlyArgValTyGlySerGluGly-----LeuSe 289
 QY 333 TCGTGGCGTCACCGCTCACAGGGGAGGGATTCCTCGGACAGTAAATGCCG---AT 389
 Db 289 rProThrLeuThrMetGlnGlyLysPheProLysIleLeuIleProGluPr 309
 QY 390 GTACGGTTTCCGATATCTGCAAAATGCCAAGTGGGGAGTGACCCAA 438
 Db 309 rIleGlnPheLeuLysValArgGluAlaThrLysLysGlyTyAlaGln 325

RESULT 13

T31093
 probable protein-tyrosine-phosphatase (EC 3.1.3.48) - medicinal leech
 N:Alternate names: receptor tyrosine phosphatase
 C:Species: Hirudo medicinalis (medicinal leech)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T31093
 R:Gershon, T.R.; Baker, M.; Nitabach, M.; Wu, P.; Macagno, E.R.
 submitted to the EMBL Data Library, December 1997
 A:Description: Two receptor tyrosine phosphatases expressed by neurons and muscle cell
 A:Reference number: Z20976
 A:Accession: T31093
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1437 <GER>
 A:Cross-references: EMBL:AF017084; NID:g2695656; PID:g2695657; PIDN:AAB91461.1
 C:Genetics:
 A:Gene: IAR1
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homolog
 ogy
 C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembra

Alignment Scores:
 Pred. No.: 17.3 Length: 1437
 Score: 74.50 Matches: 37

QY 331 ACTGCTGCCGTCACCGCTCACAGCGGAGGGATTGGCTCGG 375
 Db 127 ThrProAlaThrProAlaSerGluSerGlyIleValProGln 141

RESULT 11

F96660
 protein F2K11.10 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F96660
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Venter, J.C.; Davis, R.W.
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: F96660
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-635 <STO>
 A:Cross-references: GB:AE005173; NID:g6633854; PIDN:AAF19713.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F2K11.10
 A:Map position: 1

Alignment Scores:
 Pred. No.: 13.8 Length: 635
 Score: 75.00 Matches: 20
 Percent Similarity: 46.03% Conservative: 9
 Best Local Similarity: 31.75% Mismatches: 28
 Query Match: 9.60% Indels: 6
 DB: 2 Gaps: 1

US-09-856-221-1 (1-449) x F96660 (1-635)

QY 449 AACTGTATCAGTGGTGC-----ACTCCACCTGGCATTTTCAGA 408
 Db 12 AsnCysTyGlnPheThrAsnProAspLeuAsnThrProGluSerGluGlnSerAsn 31
 QY 407 ATAATCGGAACAGGTACATCGCATCTACTGTCGAGGCAATCCCTCCGCTGTGAG 348
 Db 32 ValIleGlySerSerIleThrSerValProValAsnAspGlyProValProLeuGlu 51
 QY 347 GCGTGTACGCGCAGCACTAAGCAGTACGAGGATCTGCTGGTGTGCGCATAGATGGTAAAT 288
 Db 52 LeuAspSerAlaAlaValSerThrSerThrSerProValGlnAlaLeuGlyHis 71
 QY 287 GACAACGGC 279
 Db 72 AspSerGly 74

RESULT 12

A95155
 hypothetical protein SPI336 [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C:Accession: A95155
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
 son, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: A95155
 A:Status: preliminary
 A:Molecule type: DNA

Percent Similarity: 41.09% Conservative: 16
Best Local Similarity: 28.68% Mismatches: 49
Query Match: 9.39% Indels: 27
DB: 2 Gaps: 9

US-09-856-221-1 (1-449) x T19050 (1-1437)

QY 55 TGTCAGAACCCGTTAGAGAGCCGAGTCAACACCTTCGTATCATATATCAGCAT 114
Db 542 TrpLysAlaProArg-----LysAlaGlyLeuSerArgTyrGlnLeuGluGly 559
QY 115 AAAATCGGCAACTCGTCAACGCGGCTTCGCGGGAACACTACTGCAAAATTCGTTA 174
Db 560 ArgLysArgValSerSerAspHisThrGlyLeu-----GluGluThrValIleAspThr 577
QY 175 ACCGCTTTCTTCTCCTCAGATAAACAACAACTGCAAACTTAC-----TGGCAGAG 228
Db 578 LysLysSerPhe-----GlnValGluLysGlyAlaGlnSerPheArgIleGluAsnLeu 595
QY 229 TTAGCAACAGCGCTATATACTTACGTCTATATCTGACA-----AsnIleThrAlaLysPheIleAspGlyIle 613
Db 596 LeuSerAsnThrLeuTyrSerIle-----AsnIleThrAlaLysPheIleAspGlyIle 613
QY 271 GATGTCAGCCCTTGTCAATACCATCTATCGACAGCAGATCGTCCGTACTGCTT 330
Db 614 ThrGlyProPtyrGlnLeuArgValGluThrSerProSerAspProProValLeu--- 632
QY 331 AGTGCTGCGCTCACCGCTTCACAGCGGAGGGGAT----- 366
Db 633 ---GluAlaProThrLeuLeuArgValThrGlyAspLeuLeuValMetArgLeuSer 651
QY 367 TTGCTCGGACAGTAATCGCGATGATC 393
Db 652 LeuPro---SerValGlnProMetTyr 659

RESULT 14

T19050
hypothetical protein C07E3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19050
R:Matthews, P.
submitted to the EMBL Data Library, June 1995
A:Reference number: 219056
A:Accession: T19050
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1186 <EIL>
A:Cross-references: EMBL:249908; PIDN:CAA90094.1; GSPDB:GN00020; CESP:C07E3.3
A:Experimental source: clone C07E3
C:Genetics:
A:Gene: CESP:C07E3.3
A:Map position: 2
A:Introns: 36/3; 79/3; 1128/3; 1160/3

Alignment Scores:
Pred. No.: 19.1 Length: 1186
Score: 74.00 Matches: 37
Percent Similarity: 43.51% Conservative: 20
Best Local Similarity: 28.24% Mismatches: 39
Query Match: 9.33% Indels: 35
DB: 2 Gaps: 7

US-09-856-221-1 (1-449) x T19050 (1-1186)

QY 63 ACCCGCTTAGA-----AGAGCGCGCAGTCAACCATTCGTGATCATATATCA 110
Db 666 ThrProLeuGlnHisTyrIleTyrArgSerArgSerSer-----ThrTyrSer 682
QY 111 GCATAAATCGCGCAACTCGCTCAACCGCGGCTTTCGCGGGAACGTAAGTCAAAATTC 170
Db 683 ValProProValGluGluAlaSerThrArgThrValAlaSerPro----- 697

QY 171 GTTAACCGCTTTGTCTCT-----TCCTCAGATAAACAACAACTGCATAAGTTACTG 221
Db 698 ---SerArgTyrThrProProSerSerSerProLysMetHisLeuThrHisLeuGlu 716
QY 222 GCAGAGCTTAGCACAACCGCTATATACTTACGTATATCTGACAATTCAGTGCACGCC 281
Db 717 ArgGluValProGlyThrProLeu-Asn- - - - -HisTyrIleSer-----ArgPr 731
QY 282 GTTGTCATTACCAATCTATCGACACACAGAT----- 315
Db 731 oArgSerThrProIleTyrSerThrProGluAspArgAsnValValSerProSerAr 751
QY 316 -----CGTTCGCTACTGTAGTCTGCGTACCGCTCACACAGCGCGAGG 362
Db 751 gTyrThrProProProSerValThrMetAlaLysLeuHisLeuThrProLeuGlnPr 771
QY 363 GGATTTGCTCGACAGATGATCGGATGATC 393
Db 771 oGluValProSerThrProLeuGlnHisTyr 781

RESULT 15

I51648
transcription initiation factor IID - African clawed frog
N:Alternate names: TATA-binding protein
C:Species: Xenopus laevis (African clawed frog)
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 03-Mar-2000
C:Accession: I51648; S68859; S26292
R:Hashimoto, S.; Fujita, H.; Hasegawa, S.; Roeder, R.G.; Horikoshi, M.
Nucleic Acids Res. 20, 3788, 1992
A:Title: Conserved structural motifs within the N-terminal domain of TFIID tau from X
A:Reference number: I51648; MUID:92350691; PMID:1641350
A:Accession: I51648
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-297 <HAS>
A:Cross-references: EMBL:X66033; NID:965148; PIDN:CAA46832.1; PID:965149
R:Labhart, P.
FEBS Lett. 386, 110-114, 1996
A:Title: Phosphorylation of the N-terminal domain of Xenopus TATA-box binding protein
A:Reference number: S68859; MUID:96228045; PMID:8647263
A:Accession: S68859
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-54, 'R', 56-176, 'I', 178-297 <IAB>
A:Note: Only a list of differences from sequence S26292 is shown in this paper
C:Genetics:
A:Gene: TFIIDtau
C:Superfamily: human transcription initiation factor IID
C:Keywords: DNA binding; nucleus; transcription initiation

Alignment Scores:
Pred. No.: 18 Length: 297
Score: 73.50 Matches: 25
Percent Similarity: 48.15% Conservative: 14
Best Local Similarity: 30.86% Mismatches: 33
Query Match: 9.27% Indels: 9
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x I51648 (1-297)

QY 157 CGTACTGCAAAATCGTTAAACCGCTTTGTTCTCTCCTCAGATAAACAACAACTGCAAAAGT 216
Db 43 GlnThrThrAsnSerLeuSerIleLeu---GluGluGlnGlnArgGlnGlnThr 61
QY 217 TACTGGCAGACGTTAGCACAA-----CGCTATATTAACCTACGTCAT 258
Db 62 GlnGlnSerThrLeuGlnGlnGlyAsnGlnGlySerGlyGlnThrProGlnLeuPheHis 81
QY 259 AATCTGACAATTCATGGTCAGCGCTG-----TCATTACCACTATTCGACACACAGCA 312
Db 82 ProGlnThrLeuThrAlaProLeuProGlyAsnThrProLeuTyrProSerProIle 101

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:45:23 ; Search time 10.6966 Seconds
(without alignments)
1826.456 Million cell updates/sec

Title: us-09-856-221-4
Perfect score: 605
Sequence: 1 gtaagaggcagttatgcaaa.....atttggaanaagatgaacgg 332

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlp
-MODEL-frame-n2p.model -DEV-xlp
-Q/cgn2_1/USPTO.spool/US09856221/runat_15012003_153926_12521/app_query.fasta_1.1948
-DB-issued_Patents_AA -QMT-fastan -SUFFIX-rai -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cd1
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-US09856221.ecgn_1.1.24.@runat_15012003_153926_12521 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	440	72.7	2522	4	US-09-251-645-13
2	70.5	11.7	507	1	US-08-363-475-19
3	70.5	11.7	532	1	US-08-363-475-22
4	67.5	11.2	680	2	US-08-674-351-2
5	67.5	11.2	707	4	US-09-021-560-4
6	67.5	11.2	744	4	US-09-021-560-2
7	64	10.6	218	4	US-09-081-689-6
8	64	10.6	218	4	US-09-073-541A-16
9	64	10.6	218	4	US-09-305-984-16
10	64	10.6	464	4	US-09-604-978-7
11	64	10.6	919	4	US-08-377-503-2
12	64	10.6	919	4	US-08-178-019-2

13	63	10.4	469	4	US-09-131-750-29	Sequence 29, Appl
14	61.5	10.2	921	1	US-09-404-670-4	Sequence 4, Appl
15	60	9.9	921	1	US-07-718-575-14	Sequence 14, Appl
16	60	9.9	921	1	US-08-481-206-14	Sequence 14, Appl
17	60	9.9	921	2	US-08-486-269A-14	Sequence 14, Appl
18	59.5	9.8	154	4	US-09-134-001C-2992	Sequence 2992, Ap
19	59.5	9.8	503	4	US-09-562-737-63	Sequence 63, Appl
20	58.5	9.7	430	4	US-09-182-816-28	Sequence 28, Appl
21	58.5	9.7	430	4	US-09-471-528-28	Sequence 28, Appl
22	58.5	9.7	430	4	US-09-634-530-28	Sequence 28, Appl
23	58.5	9.7	456	4	US-09-134-001C-3771	Sequence 3771, Ap
24	58.5	9.7	465	4	US-09-182-816-23	Sequence 23, Appl
25	58.5	9.7	465	4	US-09-471-528-23	Sequence 23, Appl
26	58.5	9.7	465	4	US-09-634-530-23	Sequence 23, Appl
27	58	9.6	484	2	US-08-252-493C-9	Sequence 9, Appl
28	58	9.6	484	3	US-09-276-197-9	Sequence 9, Appl
29	58	9.6	497	4	US-09-457-046B-65	Sequence 65, Appl
30	58	9.9	1165	1	US-08-240-357-2	Sequence 2, Appl
31	58	9.6	2233	2	US-08-569-853-1	Sequence 1, Appl
32	58	9.6	2233	3	US-08-987-439-1	Sequence 1, Appl
33	57.5	9.5	286	4	US-09-172-952-17	Sequence 17, Appl
34	57.5	9.8	440	4	US-08-759-628-4	Sequence 4, Appl
35	57.5	9.5	907	3	US-08-863-102-2	Sequence 2, Appl
36	57	9.4	222	4	US-09-134-001C-4748	Sequence 4748, Ap
37	57	9.4	741	4	US-09-001-984C-106	Sequence 106, App
38	56.5	9.3	119	1	US-08-340-539A-16	Sequence 16, Appl
39	56.5	9.3	342	4	US-08-096-181A-12	Sequence 12, Appl
40	56.5	9.3	342	4	US-08-096-181A-14	Sequence 14, Appl
41	56.5	9.3	342	5	PCT-US94-08326-12	Sequence 12, Appl
42	56.5	9.3	342	5	PCT-US94-08326-14	Sequence 14, Appl
43	56.5	9.3	361	4	US-08-096-181A-8	Sequence 8, Appl
44	56.5	9.3	361	5	PCT-US94-08326-8	Sequence 8, Appl
45	56.5	9.3	363	4	US-08-096-181A-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-251-645-13
; Sequence 13, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-13

Alignment Scores:			
Pred. No.:	8.2e-53	Length:	2522
Score:	440.00	Matches:	84
Percent Similarity:	88.18%	Conservative:	13
Best Local Similarity:	76.36%	Mismatches:	13
Query Match:	72.73%	Indels:	0
DB:	4	Gaps:	0

US-09-856-221-4 (1-332) x US-09-251-645-13 (1-2522)

QY 3 GAACGGCACTATTGCAAAAACCTATCTGGAACCCACAGGCACAACTCAGGCACAG 62

Interference?

= 6281413

Db 2224 GluAlaAlaValLeuGlnLysThrSerLeuLysThrGlnGlnThrGlnAlaGln 2243
QY 63 CTGGCTTCCTACAAACAAATTCACCAATACAGCTGTGTATACAGCTGTACGTGGCGCA 122
Db 2244 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2263
QY 123 TTGGCGGTATTTATATACAGTTTATGACTTGGCTGTTCCTCTGTCTTTGATGGCTGAA 182
Db 2264 LeuAlaAlaLeuTyrPheGlnPheTyrAspLeuAlaAlaLeuAlaArgCysLeuMetAlaGlu 2283
QY 183 CAAACTACCACTATGATTAAGTGAACGATAAGCTGTACCTTCATTAAGCCCGCTGGCTGG 242
Db 2284 GlnAlaTyrArgTrpGluLeuSerAspSerAlaArgPheIleLysProGlyAlaIleTrp 2303
QY 243 CATGGCACTATGCTGTGTGTGTAGCAGGTGAACCTGTAGCTGCAATTTGGCACAGATG 302
Db 2304 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2323
QY 303 GAAAAAATTTTGGAAAAAGATGAACGG 332
Db 2324 GluAspAlaHisLeuArgArgAspLysArg 2333

RESULT 2
US-08-363-475-19
; Sequence 19, Application US/08363475
; Patent No. 5516679

GENERAL INFORMATION:
; APPLICANT: Chiang, Shu-Jen
; APPLICANT: Burnett Jr., William V.
; APPLICANT: Tonzi, Sean M.
; TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
; TITLE OF INVENTION: FUSARIUM OXYSPORUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas R. Savitsky
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,475
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitsky, Thomas R.
; REGISTRATION NUMBER: 31,661
; REFERENCE/DOCKET NUMBER: ON-0134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252 4956
; TELEFAX: (609) 252 4526
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 507 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-363-475-19

Alignment Scores:
Pred. No.: 0.33
Score: 70.50
Matches: 31
Percent Similarity: 37.25%
Conservative: 26
Mismatches: 45
Best local Similarity: 20.28%
Query Match: 11.65%
Indels: 51
Gaps: 5
DB:

US-09-856-221-4 (1-332) x US-08-363-475-19 (1-507)
QY 3 GAAGCGCAGTATTGCAAAAAAACTATCTGAAACCAACAGGCACAAACTCAGGCACAG 62
Db 259 GluAlaLeuIleCysGlyLysGly-----GlnThrGluAsnCysLeuThrLysAlaLys 276
QY 63 CTGGCTTCCTACAAAGCAATTCAGC----- 89
Db 277 IleGluThrValArgLysValPheSerProLeuTyrThrThrAsnGluThrTyrValTyr 296
QY 90 -----AATACAGCGTGTATATACTGGCTACGTGGCGCGATTGGCGGCT 131
Db 297 ProArgAlaValProGlyAlaAsnAlaLeuPheAsnPheValValAlaGluThrProPhe 316
QY 132 ATTTAT-----TATCAGTTT----- 146
Db 317 ValTyrSerThrGluTrpTyrGlnTyrValIleTrpGluAspProGluTrpAsnProAsp 336
QY 147 -----TATGACTTGGCTGTTTCCCTGTGTGTGTGATGGCTGAACAACT 188
Db 337 ThrIleGlyProLysAspTyrArgGlyAlaGluMetAsnProTyrAspIleGluThr 356
QY 189 TACCAGTATGAATTTGAACCATAAAGCTGTACGCTTCATTAAAGCCCGTGGCGCATGGC 248
Db 357 TrpGluGlyAspLeuSerLysPheArgLysArgGlyAsnLysMetIleHisTrpHisGly 376
QY 249 ACTTATGCTGCTGTTTGTACAGGTGAA----- 275
Db 377 LeuGlnAspGlyLeuIleSerAlaGluAsnSerAspTyrTyrAsnHisValSerArg 396
QY 276 ACCTTGATGCTGAATTTGGCAGACAGATGGAAAAAACTAT 314
Db 397 ThrMetGlyLeuAsnSerSerGlnLeuAspGlnPheTyr 409

RESULT 3
US-08-363-475-22
; Sequence 22, Application US/08363475
; Patent No. 5516679
GENERAL INFORMATION:
; APPLICANT: Chiang, Shu-Jen
; APPLICANT: Burnett Jr., William V.
; APPLICANT: Tonzi, Sean M.
; TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
; TITLE OF INVENTION: FUSARIUM OXYSPORUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas R. Savitsky
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,475
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitsky, Thomas R.
; REGISTRATION NUMBER: 31,661
; REFERENCE/DOCKET NUMBER: ON-0134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252 4956
; TELEFAX: (609) 252 4526
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear


```
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 825-144P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 707 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-021-560-4
Alignment Scores:
Pred. No.: 0.992 Length: 707
Score: 67.50 Matches: 27
Percent Similarity: 35.79% Conservative: 7
Best Local Similarity: 28.42% Mismatches: 32
Query Match: 11.16% Indels: 29
DB: 4 Gaps: 3
US-09-856-221-4 (1-332) x US-09-021-560-4 (1-707)
QY 66 GCCTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATACCTGGCTACGTCGGCGATTG 125
Db 562 AlaPheIleLysSerSerPhePheAsnSerAlaSerAspValTrpThrTyrGlyPheGly 581
QY 126 GCGGCTATTTATATACAGTTTATGACTGGCTGTTCCCTGTTGTGATGGCTGACAA 185
Db 582 AlaAspAlaLeuTyrAsnPhe----- 588
QY 186 ACTTACCAGTATGAATGAACGATAAGCTGTACGCTTCATTAAAGCCGCGTGCCTGGCAT 245
Db 589 -----IleAsnAspLysAlaThrAsnPheLeu-----GlyLysAsnAsn 601
QY 246 GGCACCTATGCTGTTGTTAGCAGTGAACCTTG----- 281
Db 602 LysLeuSerValGlyLeuPheGlyGlyIleAlaLeuAlaGlyThrSerTrpLeuAsnSer 621
QY 282 ---ATGCTGAATTTGGCAGATGGAAAAAACTATTTGGAAAAA 323
Db 622 GluTyrValAsnLeuAlaThrMetAsnAsnValTyrAsnAlaLys 636
RESULT 6
US-09-021-560-2
; Sequence 2, Application US/09021560
; Patent No. 6410719
; GENERAL INFORMATION:
; APPLICANT: BOREN, THOMAS
; APPLICANT: NORMARK, STAFFAN
; APPLICANT: ARNOVIST, ANNA
; APPLICANT: ILVER, DAG
; TITLE OF INVENTION: BLOOD GROUP ANTIGEN BINDING PROTEIN AND
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,560
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 825-144P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-021-560-2
Alignment Scores:
Pred. No.: 1.01 Length: 744
Score: 67.50 Matches: 27
Percent Similarity: 35.79% Conservative: 7
Best Local Similarity: 28.42% Mismatches: 32
Query Match: 11.16% Indels: 29
DB: 4 Gaps: 3
US-09-856-221-4 (1-332) x US-09-021-560-2 (1-744)
QY 66 GCCTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATACCTGGCTACGTCGGCGATTG 125
Db 599 AlaPheIleLysSerSerPhePheAsnSerAlaSerAspValTrpThrTyrGlyPheGly 618
QY 126 GCGGCTATTTATATACAGTTTATGACTGGCTGTTCCCTGTTGTGATGGCTGACAA 185
Db 619 AlaAspAlaLeuTyrAsnPhe----- 625
QY 186 ACTTACCAGTATGAATGAACGATAAAGCTGTACGCTTCATTAAAGCCGCGTGCCTGGCAT 245
Db 626 -----IleAsnAspLysAlaThrAsnPheLeu-----GlyLysAsnAsn 638
QY 246 GGCACCTATGCTGTTGTTAGCAGTGAACCTTG----- 281
Db 639 LysLeuSerValGlyLeuPheGlyGlyIleAlaLeuAlaGlyThrSerTrpLeuAsnSer 658
QY 282 ---ATGCTGAATTTGGCAGATGGAAAAAACTATTTGGAAAAA 323
Db 659 GluTyrValAsnLeuAlaThrMetAsnAsnValTyrAsnAlaLys 673
RESULT 7
US-09-081-689-6
; Sequence 6, Application US/09081689
; Patent No. 6165992
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; APPLICANT: Zalacain, Madgalenaa
; APPLICANT: Throup, John
; APPLICANT: Biswas, Sanjoy
; TITLE OF INVENTION: Histidine Kinase
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,689
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,347
```



```

Db 150 AlaLysGluLeuGluLeuLeuAspTyrLeuValLysAsnGlu 163
      :::::::::: :::::::::: :::::::::: ::::::::::
RESULT 10
US-09-604-978-7
; Sequence 7, Application US/09604978
; Patent No. 6455674
; GENERAL INFORMATION:
; APPLICANT: Eibat, Paz
; TITLE OF INVENTION: HYPoxIA-REGULATED GENES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6455674thwestern Hwy., Suite 401
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/604,978
; FILING DATE: 28-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/138,112
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 0168.00034
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 5395055
; INFORMATION FOR SEQ ID NO: 7:
; LENGTH: 464 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-604-978-7
Alignment Scores:
Pred. No.: 2.61 Length: 464
Score: 64.00 Matches: 29
Percent Similarity: 33.59% Conservative: 15
Best Local Similarity: 22.14% Mismatches: 45
Query Match: 10.58% Indels: 42
DB: 4 Gaps: 3
US-09-856-221-4 (1-332) x US-09-604-978-7 (1-464)
QY 12 GTATTGCAAAAACTATCTGGAACCCACAGCCACAACTCAGCAGCTGCCCTTC 71
      :::::::::: :::::::::: :::::::::: ::::::::::
Db 17 LeuLeuArgSerGlyGlnGlyGluGlnGlnGlnGlnGlnAlaGlnArgCysPhe 36
      :::::::::: :::::::::: :::::::::: ::::::::::
QY 72 CTACAA-----AGCAATTC 86
      :::::::::: ::::::::::
Db 37 CysGlnValSerGlyTyrLeuAspCysThrCysAspValGluThrIleAspLysPhe 56
      :::::::::: ::::::::::
QY 87 ACAATACAGCGTTGTATACCTGCTACGCTGGCGGCTATTTATTATCAGTTT 146
      :::::::::: ::::::::::
Db 57 AsnAsnTyrArgLeuPheProArgLeuGlnLysLeuLeuGluSerAspTyrPheArgTyr 76
      :::::::::: ::::::::::
QY 147 TAT----- 149
      ::::::::::

```

```

Db 77 TyrLysValAsnLeuArgLysProCysProPheTrpAsnAspIleAsnGlnCysGlyArg 96
      :::::::::: :::::::::: :::::::::: ::::::::::
QY 150 ---GACTGGCTGTTCCCTGCTGTTGATGGCTGAA-----CAA 185
      :::::::::: :::::::::: ::::::::::
Db 97 ArgAspCysAlaValLysProCysHisSerAspGluValProAspGlyIleLysSerAla 116
      :::::::::: :::::::::: ::::::::::
QY 186 ACTTACCAAGTATGAATTAACGCTGACGCTTCATTAAAGCCGCTGCTGGCAT 245
      :::::::::: :::::::::: ::::::::::
Db 117 SerTyrLysTyrSerLysGluAlaAsnLeuLeuGluCysGluProAlaGluArgLeu 136
      :::::::::: :::::::::: ::::::::::
QY 246 GCACCTATGCTGTTGTTAGCAGGTGAAACC 278
      :::::::::: ::::::::::
Db 137 GlyAlaValAspGluSerLeuSerGluGluThr 147
      ::::::::::
RESULT 11
US-08-377-503-2
; Sequence 2, Application US/08377503
; Patent No. 6322999
; GENERAL INFORMATION:
; APPLICANT: Kamboj, Rajender
; APPLICANT: Elliot, Candace E.
; TITLE OF INVENTION: Kainate-Binding, Human CNS Receptors of
; TITLE OF INVENTION: the EAA5 Family
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street
; CITY: N.W.
; STATE: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/377,503
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,210
; FILING DATE: 17-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/192
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 919 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-377-503-2
Alignment Scores:
Pred. No.: 3.41 Length: 919
Score: 64.00 Matches: 22
Percent Similarity: 40.18% Conservative: 23
Best Local Similarity: 19.64% Mismatches: 61
Query Match: 10.58% Indels: 6
DB: 4 Gaps: 1
US-09-856-221-4 (1-332) x US-08-377-503-2 (1-919)
QY 6 GCGCAGTATGCAAAAACTATCTGGAACCCACAGCCACAACTCAGCAGCTG 65
      :::::::::: :::::::::: ::::::::::
Db 293 SerAlaIleValGluLysTrpSerMetGluArgLeuGlnAlaProArgAlaGluSer 312
      :::::::::: ::::::::::

```


Search completed: January 15, 2003, 15:57:51
Job time : 14.6966 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:46:43 ; Search time 6.22838 Seconds
(without alignments)
2119.222 Million cell updates/sec

Title: US-09-856-221-4
Perfect score: 605
Sequence: 1 gtaagcgcagctattgcaaa.....atttgaaaagaatgaacg 332

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-DB=/cgn2_1/USPTO.spool/US09856221/runat_15012003.153926.12555/app-query.fasta_1.1948
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09856221.ecgn_1_14/runat_15012003.153926.12555
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	446	73.7	2516	10	US-09-817-514A-2
2	418	69.1	2504	10	US-09-817-514A-8
3	68	11.2	696	10	US-09-815-242-5443
4	68	11.2	698	10	US-09-815-242-12325

Alignment Scores:
Pred. No.: 5.98e-49
Score: 446.00
Percent Similarity: 89.09%
Best Local Similarity: 78.18%

Length: 2516
Matches: 86
Conservative: 12
Mismatches: 12

ALIGNMENTS

RESULT 1
US-09-817-514A-2
; Sequence 2, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: french-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2

312	10	US-09-886-055-171	Sequence 171, Appl
218	10	US-09-737-068-6	Sequence 6, Appli
698	9	US-09-801-220-4	Sequence 4, Appli
708	8	US-08-834-666A-2	Sequence 2, Appli
489	10	US-09-815-242-13932	Sequence 13932, A
733	8	US-08-834-666A-4	Sequence 4, Appli
745	8	US-08-834-666A-6	Sequence 6, Appli
47	10	US-09-864-761-37729	Sequence 37729, A
359	10	US-09-764-864-986	Sequence 986, App
487	10	US-09-815-242-12036	Sequence 12036, A
163	10	US-09-861-451A-80	Sequence 80, Appl
323	10	US-09-815-242-10578	Sequence 10578, A
406	9	US-09-738-626-4686	Sequence 4686, Ap
777	10	US-09-864-761-37995	Sequence 37995, A
777	10	US-09-864-761-46740	Sequence 46740, A
159	10	US-09-811-284-168	Sequence 168, App
207	9	US-09-764-868-981	Sequence 981, App
165	10	US-09-864-761-34905	Sequence 34905, A
489	10	US-09-815-242-10215	Sequence 10215, A
124	10	US-09-811-284-165	Sequence 165, App
497	9	US-09-866-570A-65	Sequence 65, Appl
574	10	US-09-953-688A-1	Sequence 1, Appli
597	9	US-10-025-222A-40	Sequence 40, Appl
368	10	US-09-801-368-308	Sequence 308, App
426	10	US-09-764-853-667	Sequence 667, App
442	9	US-09-870-759-55	Sequence 55, Appl
716	9	US-10-008-355-7	Sequence 7, Appli
261	10	US-09-764-864-961	Sequence 961, App
419	9	US-09-738-626-4134	Sequence 4134, Ap
429	10	US-09-815-242-5541	Sequence 5541, Ap
430	10	US-09-815-242-12226	Sequence 12226, A
430	10	US-09-815-242-12825	Sequence 12825, A
93	430	US-09-815-242-13050	Sequence 13050, A
683	9	US-09-738-626-6961	Sequence 6961, Ap
235	10	US-09-799-777-67	Sequence 67, Appl
265	10	US-09-815-242-10621	Sequence 10621, A
93	265	US-09-815-242-10621	Sequence 10621, A
270	10	US-09-841-132-404	Sequence 404, Appl
291	10	US-09-829-631A-10	Sequence 10, Appl
394	9	US-09-738-626-4273	Sequence 4273, Ap
437	10	US-09-829-631A-8	Sequence 8, Appli

Reference

```
Query Match: 73.72% Indels: 0
DB: 10 Gaps: 0
US-09-856-221-4 (1-332) x US-09-817-514A-2 (1-2516)
QY 3 GAAGCGCAGTATTGCAAAACAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
Db 2218 GluAlaLaValLeuGlnTyThrSerLeuLysThrGlnGlnGluGlnThrGlnSerGln 2237
QY 63 CTGGCTTCTTACAAACCAATTCAGCAATACAGCGTTGTATACACTGGCTAGCTGGCGCA 122
Db 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrsnTrpLeuArgGlyArg 2257
QY 123 TTGGCGGCTATTATATCATCTAGTTTATGACTTGGCTGTTCCTCGTGTGTGATGCTGAA 182
Db 2258 LeuAlaAlaIleTyThrPheGlnPheTyAspLeuAlaValAlaArgCysLeuMetAlaGlu 2277
QY 183 CAACTTACCAAGTAAATGAAGCAATGAAGCTGTACGCTTCAATGAAGCCGCTGCTGG 242
Db 2278 GlnAlaTyArgTrpGluLeuAsnAspSerAlaArgPheLeuLysProGlyAlaTrp 2297
QY 243 CATGCCACTTATGCTGCTTTGTAGCAGGTGAACCTTGTATGCTGAATTTGGCAGATG 302
Db 2298 GlnGlyThyTyAlaGlyLeuAlaGlyGluThyLeuMetLeuSerLeuAlaGlnMet 2317
QY 303 GAAAAAACTATTGGAAGAAGATGAACGG 332
Db 2318 GluAspAlaHisLeuLysArgAspLysArg 2327
RESULT 2
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Alignment Scores:
Pred. No.: 2,37e-45 Length: 2504
Score: 418.00 Matches: 78
Percent Similarity: 84.55% Conservative: 15
Best Local Similarity: 70.91% Mismatches: 17
Query Match: 69.09% Indels: 0
DB: 10 Gaps: 0
US-09-856-221-4 (1-332) x US-09-817-514A-8 (1-2504)
QY 3 GAAGCGCAGTATTGCAAAACAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
Db 2212 GluAlaAlaGluMetGlnLysGluTyLeuLysThrGlnGlnAlaGlnAlaGln 2231
QY 63 CTGGCTTCTTACAAACCAATTCAGCAATACAGCGTTGTATACACTGGCTAGCTGGCGCA 122
Db 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrsnTrpLeuArgGlyArg 2251
QY 123 TTGGCGGCTATTATATCATCTAGTTTATGACTTGGCTGTTCCTCGTGTGTGATGCTGAA 182
Db 2252 LeuSerGlyIleTyThrPheGlnPheTyAspLeuAlaValSerArgCysLeuMetAlaGlu 2271
QY 183 CAAACTTACCAAGTAAATGAAGCAATGAAGCTGTACGCTTCAATGAAGCCGCTGCTGG 242
Db 2272 GlnSerTyGlnTrpGluAlaAsnAspSerIleSerPheValLysProGlyAlaTrp 2291
QY 243 CATGCCACTTATGCTGCTTTGTAGCAGGTGAACCTTGTATGCTGAATTTGGCAGATG 302
Db 639 TyrGlyAlaPheValGlyLeuPheProGlyLysAspAlaLeuHisIleSerGlnIle 658
QY 303 GAAAAAACTATTGGAAGAAGATGAACGG 332
Db 627 GlnThrTyrgln-----AlaThrValLysArgIleGluLys----- 638
US-09-815-242-5443
; Sequence 5443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELIPIRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5443
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5443
Alignment Scores:
Pred. No.: 1.6 Length: 696
Score: 68.00 Matches: 17
Percent Similarity: 57.14% Conservative: 11
Best Local Similarity: 34.69% Mismatches: 13
Query Match: 11.24% Indels: 8
DB: 10 Gaps: 2
US-09-856-221-4 (1-332) x US-09-815-242-5443 (1-696)
QY 183 CAAACTTACCAAGTAAATGAAGCAATGAAGCTGTACGCTTCAATGAAGCCGCTGCTGG 242
Db 627 GlnThrTyrgln-----AlaThrValLysArgIleGluLys----- 638
QY 243 CATGCCACTTATGCTGCTTTGTAGCAGGTGAACCTTGTATGCTGAATTTGGCAGATG 302
Db 639 TyrGlyAlaPheValGlyLeuPheProGlyLysAspAlaLeuHisIleSerGlnIle 658
QY 303 GAAAAAACTATTGGAAGAAGATGAACGG 332
Db 627 GlnThrTyrgln-----AlaThrValLysArgIleGluLys----- 638
```

```
Db 659 SerLysAsnArgIleGluLysValGlu 667
RESULT 4
US-09-815-242-12325
; Sequence 12325, Application US/09815242
; Patent No. US20020061369A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12325
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12325
Alignment Scores:
Pred. No.: 1,6 Length: 698
Score: 68.00 Matches: 17
Percent Similarity: 57.14% Conservative: 11
Best Local Similarity: 34.69% Mismatches: 13
Query Match: 11.24% Indels: 8
DB: 2
US-09-856-221-4 (1-332) x US-09-815-242-12325 (1-698)
QY 183 CAACCTACAGTATGATTAAGCATTAAGCTTCAATTAAGCCGGTGGCTGG 242
Db 627 GlnThrTyrGln-----AlaThrValLysArgIleGluLys----- 638
QY 243 CATGGCACTATGCTGGTTTGTAGCAGGTGAACCTTGATGCTGAATTTGGCACAGATG 302
Db 639 TyrGlyAlaPheValGlyLeuPheProGlyLysAspAlaLeuLeuHisIleSerGlnIle 658
QY 303 GAAAAAAGCTATTTCGAAAAAGATGAA 329
Db 659 SerLysAsnArgIleGluLysValGlu 667
RESULT 5
US-09-886-055-171
; Sequence 171, Application US/09886055
; Patent No. US20020132273A1
; GENERAL INFORMATION:
; APPLICANT: STRYER, LUBERT
; APPLICANT: ZOZULYA, SERGEY
; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
; TITLE OF INVENTION: BIOSENSORS OF CHEMICAL SENSANTS
; FILE REFERENCE: 078003-Q277150
; CURRENT APPLICATION NUMBER: US/09/886,055
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 60/213,812
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 171
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-055-171
Alignment Scores:
Pred. No.: 3,11 Length: 312
Score: 65.00 Matches: 27
Percent Similarity: 37.61% Conservative: 14
Best Local Similarity: 24.77% Mismatches: 50
Query Match: 11.07% Indels: 18
DB: 2
US-09-856-221-4 (1-332) x US-09-886-055-171 (1-312)
QY 311 GTTTTTTCCATCTGTGCCAAATTCAGCATCAAGGTTTCACCTGCTAACAACACGACATA 252
Db 57 MetTyrPhePheLeuLeuCysAsnLeuSerPheLeuGluIleTyrThrThrValIle 76
QY 251 AGTGCCATGCCAGCAGCGGGCTTAATGAAGCGTACAGCTTTATCGTTCAATTCACTACTG 192
Db 77 ProLysLeuLeuGlyThrPheValAlaAlaArgThrValIleCysMetSerCysLeu 96
QY 191 GTAAGTTTGTTCAGCCATCAACACAGGAGGAAACAGCAAGTCATAAAACG----- 141
Db 97 LeuGlnAlaPhePheHisPhePheValGlyThrGluPheLeuIleLeuThrIleMet 116
QY 140 -----ATATAAATACGCCCAATCGCCCAATCGCCCAATCGCCCAATCGCCCAAT 102
Db 117 SerPheAspArgTyrLeuThrIleCysAsnProLeuHisHisProThrIleMetThrSer 136
QY 101 CAACGCTGTAATGCTGAATTTGCTTTGTAGGAAGCGCAGCTGTGCTGTGCTGTGCTGTG 42
Db 137 LysLeuCysLeu-----GlnLeuAlaLeuSerSerTrpVal 148
QY 41 TTGGCTTCCAGATAGTTTTTTTGCAG 15
Db 149 ValGlyPheThrIleValPheCysGln 157
RESULT 6
US-09-737-068-6
; Sequence 6, Application US/09737068
; Patent No. US20010020010A1
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; APPLICANT: Zalacain, Madgalenaa
; APPLICANT: Throup, John
; APPLICANT: Biswas, Sanjoy
; TITLE OF INVENTION: Histidine Kinase
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/737,068
; FILING DATE:
```

CLASSIFICATION:
PRIOR APPLICATION DATA: 09/081,689
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-737-068-6

Alignment Scores:
Pred. No.: 3.78 Length: 218
Score: 64.00 Matches: 26
Percent Similarity: 48.94% Conservative: 20
Best Local Similarity: 27.66% Mismatches: 24
Query Match: 10.58% Indels: 24
DB: 10 Gaps: 6

US-09-856-221-4 (1-332) x US-09-737-068-6 (1-218)

QY 57 GCACAGTGGCC-----TTCCTACAAAGCAATTCACGATACAGCGTTGTATACACTGG 110
Db 91 AlaserLeuAlaAspGlyTyrLeuGluLysProPheSer-----LeuSerLeu 106
QY 111 CTACGTGGCGGATTGGCGGCTATTATATACGATTTTATGACTTGGCTGGTTCCTCGTGT 170
Db 107 LeuLysValArgValAspAlaIlePheLysArgTyrTyrAspThrGly-----122
QY 171 TTGATGGTGACAAACTTACCAGTATGAAATTCAGCATAAGCTGTACGCTTCATTAAG 230
Db 123 -----ArgIlePheSerTyrLys-----AspThrLysValAspPhe-----134
QY 231 CCGGCGTGGCAGTGGCAGCTATGCTGGTTGTTAGCAGGTGAACCTTTGATCGCTGAAT 290
Db 135 -----GlusertTyrSerAlaserLeuAlaGlyGlnGluValProIleAsn 149
QY 291 TTGGCACAGTGGAA---AAAACTATTGGAAAAAGATGAA 329
Db 150 AlaLysGluLeuGluIleLeuAspTyrLeuValLysAsnGlu 163

RESULT 7
US-09-801-220-4
Sequence 4, Application US/09801220
Patent No. US20020173020A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel A.
APPLICANT: Williamson, Mark
TITLE OF INVENTION: 26886, A NOVEL CARNITINE ACYLTRANSFERASE
FILE REFERENCE: 10448-026001
CURRENT APPLICATION NUMBER: US/09/801,220
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 60/187,456
PRIOR FILING DATE: 2000-03-07
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 698
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: consensus sequence
US-09-801-220-4

Alignment Scores:
Pred. No.: 5.23 Length: 698
Score: 64.00 Matches: 17
Percent Similarity: 55.56% Conservative: 13
Best Local Similarity: 31.48% Mismatches: 18
Query Match: 10.58% Indels: 6
DB: 9 Gaps: 2

US-09-856-221-4 (1-332) x US-09-801-220-4 (1-698)

QY 141 CAGTTTATGACTTGGCTGTTTCCCTGTGTGTTGATGCTGAACAACTTACCAGTATGAA 200
Db 129 ArgPheLysGluLeuLeuAspAlaserGluLeuProGluLeuAlaLysAsnGlu 148
QY 201 TTGACGATAAAGCT-----GTACGCTTATTAAAGCCCGGTGCTGCATGGC 248
Db 149 LysSerAspThrAlaPheLysArgLeuIleArgPheValProSerLeuSerTyrTyrGly 168
QY 249 ACTTATGCTGTTTGTAGCAGGTGAACCTTGTATGCTGAAT 290
Db 169 AlaTyr-----LeuLeuGlyGlyGlnProLeuCysMetAsn 180

RESULT 8
US-08-834-666A-2
Sequence 2, Application US/08834666A
Patent No. US20020044949A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Lissolo, Ling
APPLICANT: Tomb, Jean-Francois
APPLICANT: Miller, Charles
APPLICANT: Al-Garawi, Amal
TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
TITLE OF INVENTION: Corresponding Polynucleotide Molecules
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,666A
FILING DATE: 01-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,175
REFERENCE/DOCKET NUMBER: 06132/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1..19
OTHER INFORMATION:
US-08-834-666A-2

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 733 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...20
OTHER INFORMATION:
US-08-834-666A-4

Alignment Scores:
Pred. No.: 8.26 Length: 733
Score: 62.50 Matches: 26
Percent Similarity: 35.79% Conservative: 8
Best Local Similarity: 27.37% Mismatches: 32
Query Match: 10.33% Indels: 29
DB: 8 Gaps: 3

US-09-856-221-4 (1-332) x US-08-834-666A-4 (1-733)

QY 66 GCCTTCCTCAAAAGCAAAATTCAGCAATACAGCGTTGTATACCTGGCTACGTGGCGGATTG 125
Db 588 AlaPheIleIysSerSerPhePheAsnSerAlaSerAspValTrpThrTyrglyPheGly 607
QY 126 GCGCGTATTATTATACGTTTATGACTTGGCTGTTTCCCTGCTGTTGATGGCTGAACAA 185
Db 608 AlaAspAlaLeuTyraAsnPhe----- 614
QY 186 ACTTACCAGTATGAATTAAGCAGATAAGCTGTACGCTTCATTAAAGCGCGGTGCGTGGCAT 245
Db 615 -----IleAsnAspLysAlaThrAsnPheLeu-----GlyLysAsnAsn 627
QY 246 GGCACATTATGCTGTTTGTACAGGTGAACCTTG----- 281
Db 628 LysLeuSerLeuGlyLeuPheGlyIleAlaLeuAlaGlyThrSerTrpLeuAsnSer 647
QY 282 ---ATGCTGAATTTGGCAGATGGAAGAAAAAATTTTGGAAAAA 323
Db 648 GluTyrValAsnLeuAlaThrValAsnAsnValTyraAsnAlaLys 662

RESULT 11
US-08-834-666A-6
Sequence 6, Application US/08834666A
Patent No. US20020044949A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Lissolo, Ling
APPLICANT: Tomb, Jean-Francois
APPLICANT: Miller, Charles
TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
TITLE OF INVENTION: Corresponding Polynucleotide Molecules
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834.666A
FILING DATE: 01-APR-1997
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,175
REFERENCE/DOCKET NUMBER: 06132/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 745 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...20
OTHER INFORMATION:
US-08-834-666A-6

Alignment Scores:
Pred. No.: 8.29 Length: 745
Score: 62.50 Matches: 26
Percent Similarity: 35.79% Conservative: 8
Best Local Similarity: 27.37% Mismatches: 32
Query Match: 10.33% Indels: 29
DB: 8 Gaps: 3

US-09-856-221-4 (1-332) x US-08-834-666A-6 (1-745)

QY 66 GCCTTCCTCAAAAGCAAAATTCAGCAATACAGCGTTGTATACCTGGCTACGTGGCGGATTG 125
Db 600 AlaPheIleIysSerSerPhePheAsnSerAlaSerAspValTrpThrTyrglyPheGly 619
QY 126 GCGCGTATTATTATACGTTTATGACTTGGCTGTTTCCCTGCTGTTGATGGCTGAACAA 185
Db 620 AlaAspAlaLeuTyraAsnPhe----- 626
QY 186 ACTTACCAGTATGAATTAAGCAGATAAGCTGTACGCTTCATTAAAGCGCGGTGCGTGGCAT 245
Db 627 -----IleAsnAspLysAlaThrAsnPheLeu-----GlyLysAsnAsn 639
QY 246 GGCACATTATGCTGTTTGTACAGGTGAACCTTG----- 281
Db 640 LysLeuSerLeuGlyLeuPheGlyIleAlaLeuAlaGlyThrSerTrpLeuAsnSer 659
QY 282 ---ATGCTGAATTTGGCAGATGGAAGAAAAAATTTTGGAAAAA 323
Db 660 GluTyrValAsnLeuAlaThrValAsnAsnValTyraAsnAlaLys 674

RESULT 12
US-09-864-761-37729
Sequence 37729, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6


```
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37729
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006344.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.86
; OTHER INFORMATION: EST_HUMAN HIT: BE466313.1, EVALUAE 1.00e-11
US-09-864-761-37729

Alignment Scores:
Pred. No.: 6 Length: 47
Score: 61.00 Matches: 18
Percent Similarity: 52.50% Conservative: 3
Best Local Similarity: 45.00% Mismatches: 5
Query Match: 10.08% Indels: 14
DB: 10 Gaps: 4

US-09-856-221-4 (1-332) x US-09-864-761-37729 (1-47)
QY 165 CTGTGTTGATGCTGAACAACTTACCAGTATGAATTGAACGATAAAGCTAGCGTTC 224
||||| :|||:|||||:|||||:|||||:|||||:
Db 16 LeuCysPheLeu-----ThrTyrIstPrgLLeuAsnAspLys----- 28
QY 225 ATTAAGCCCGTGCCTGG---CATGGC-----ACTTATGCTGTTTGTAGCAGGT 272
||||| :|||:|||||:|||||:|||||:|||||:
Db 29 -----AsnLysTrpThrHisGlyAsnAsnThrTyrTrpGlyLeuLeuGluGly 45

RESULT 13
US-09-764-864-996
; Sequence 996, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 996
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-996

Alignment Scores:
Pred. No.: 10.6 Length: 359
Score: 61.00 Matches: 19
Percent Similarity: 43.86% Conservative: 6
Best Local Similarity: 33.33% Mismatches: 14
Query Match: 10.39% Indels: 18
DB: 10 Gaps: 3

US-09-856-221-4 (1-332) x US-09-764-864-996 (1-359)
QY 291 AATTGAGCATCAAGGTTTCACCTGCTAACAAACAGCATAGTCCATGCCAGCACGG 232
|||||:|||||:|||||:|||||:|||||:
Db 87 AsnSerSerSer-----LeuLeuAsnHisHisLysValHisAla----- 99
QY 231 GCTTAATGAACGCTACAGCTTTATCTTCATTCATCTACTGTTAGTTGTTGTCAGCCATCA 172
|||||:|||||:|||||:|||||:|||||:
Db 100 -----GlyLysGlnProTyrArgCysIleGluCysGlyLysPheLeuLysLysHis 116
QY 171 AACACA-----GGAAACACCAAGTCATAAA 145
|||||:
Db 117 SerThrPheIleAsnHisGlnArgIleHisSerArgGluLysProHisLys 133

RESULT 14
US-09-815-242-12036
; Sequence 12036, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12036
; LENGTH: 487
; TYPE: PRT
; GENERAL INFORMATION:
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 15, 2003, 15:44:53 ; Search time 15.3002 Seconds
(without alignments)
4172.064 Million cell updates/sec

Title: US-09-856-221-4

Perfect score: 605
Sequence: 1 gtaagcggcagattgca.....atttgaaaaaatgaacgg 332

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=cn2_1/USPTO.spool/US09856221/runat_15012003_153925_12500/app.query.fasta_1.1948
-DB=PIR.73 -QFMT=fastan -SUFFIX=ror -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856221.ecgn_1.1.95.erunat_15012003_153925_12500 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR.73.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	12.1	439	E71116	hypothetical prote
2	72	11.9	269	T44866	hypothetical prote
3	70.5	11.7	398	E71379	probable tpr prote
4	70.5	11.7	598	D71365	probable 64.7K tpr
5	70.5	11.7	1309	F82207	ATP-dependent heli
6	70	11.6	194	A21185	hypothetical prote
7	69.5	11.5	467	E81828	probable periplasm
8	69.5	11.5	1188	NDECKR	type I site-specif
9	69.5	11.5	1392	A43336	microtubule-vesicl
10	69.5	11.5	1427	S22695	reslin - human
11	69	11.4	353	S37873	hypothetical prote
12	68.5	11.3	258	S75950	hypothetical prote
13	68.5	11.3	312	T33215	hypothetical prote
14	68	11.2	503	AH0159	probable lysine-sp

15	68	11.2	698	2	E89901	polyribonucleotide
16	67.5	11.2	610	2	T22887	hypothetical prote
17	67.5	11.2	703	2	C71840	outer membrane pro
18	67.5	11.2	744	2	D71883	outer membrane pro
19	67	11.4	379	2	T11349	ubiquinol-cytochro
20	66.5	11.0	367	2	AH0688	probable bacteriop
21	66	10.9	469	2	H70626	probable nark3 pro
22	66	10.9	826	2	B96712	probable receptor
23	65.5	10.8	149	2	C69314	hypothetical prote
24	65.5	11.2	240	2	T15785	hypothetical prote
25	65.5	10.8	1353	2	JO0407	xanthine dehydroge
26	65	11.1	199	2	A97845	phosphatidylglycer
27	65	10.7	211	2	S34274	probable secretory
28	65	10.7	243	2	S74031	hypothetical prote
29	65	10.7	273	2	B83551	hypothetical prote
30	65	10.7	320	2	E71696	rare lipoprotein A
31	65	10.7	370	2	H70310	conserved hypothet
32	65	10.7	430	2	H81389	thiamin biosynthes
33	65	11.1	648	2	A71342	probable DNA misla
34	65	10.7	680	2	S26764	major coat protein
35	65	10.7	1505	2	S26765	genome polyprotein
36	65	10.7	2091	2	A97077	hypothetical prote
37	64.5	10.7	286	2	H64113	hemagglutinin nagl
38	64.5	10.7	418	2	G87469	ThiJ/Pfpo family p
39	64.5	10.7	440	2	B75044	hypothetical prote
40	64.5	10.7	637	2	S46771	hypothetical prote
41	64.5	10.7	1465	2	A70199	hypothetical prote
42	64	10.6	122	2	D83448	succinate dehydrog
43	64	10.6	218	2	B95070	DNA-binding respon
44	64	10.6	218	2	H97937	vncR, response reg
45	64	10.6	240	2	B82383	conserved hypothet

ALIGNMENTS

RESULT 1

E71116
hypothetical protein PH0700 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: E71116
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71116
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-439 <RAW>
A:Cross-references: GB:AF000003; MID:g3236130; PIDN:BAA29791.1; PID:d1030734; PID:g32
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0700

Alignment Scores:			
Pred. No.:	1.95	Length:	439
Score:	73.00	Matches:	18
Percent Similarity:	55.56%	Conservative:	12
Best Local Similarity:	33.33%	Mismatches:	20
Query Match:	12.07%	Indels:	4
DB:	2	Gaps:	2

US-09-856-221-4 (1-332) x E71116 (1-439)

QY	132	ATTATATACATTTTAT-----GACTTGGCTGTTCCCTGCTGTTGATGCGTGAA	182
Db	122	IllePheTyMetPhePheLeuAlaAlaThrIleProLeuProLeuSerLeuIleSerArg	141
QY	183	CAAACTTACCATGATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCGGTGGCTGG	242
Db	142	GlupPheArgArgTyGluIleAsnLysAlaIleGlyArgPheAsnGluLeuGlyGlyTrp	161

A:Gene: TP0316
C:Keywords: translational frameshift
F;370-371/Region: plus-one translational frameshift

Alignment Scores:
Pred. No.: 3.96 Length: 398
Score: 70.50 Matches: 37
Percent Similarity: 35.86% Conservative: 15
Best Local Similarity: 25.52% Mismatches: 35
Query Match: 11.65% Indels: 58
DB: 2 Gaps: 6

US-09-856-221-4 (1-332) x E71379 (1-398)

QY 53 TCAGGCAACGGCTGGCTCCCTACAAAGCAAATTCAGCAATACAGCGTGTGTATAACTGGCT 112
Db 20 SerGlyThrAlaGlyValLeuThrProGlnVal-SerGlyThrAlaGlnLeuGlnTrp-- 38
QY 113 ACGTGGCGATTGGCGGCTATTTATTATCAG----- 143
Db 39 -----GlyIleAlaPheGlnLysAsnProArgThrGlyProGlyLysHI 53
QY 144 -----TTTTATGACTTGGCTGTTCCTGTTGATGCCTGAACA 184
Db 53 sThrHisGlyPheAthrThrAsnSerLeuThriIeSerLeuProLeuValSerLysHI 73
QY 185 AACTTAC-----CAGTATGATTCGAACA 208
Db 73 sThrHisThrArgArgGlyGluAlaArgSerGlyValTrpAlaGlnLeuLysAs 93
QY 209 TAAAGCTGTACGC-----TTCATTAAGCCCGG 235
Db 93 pLeuAlaValGluLeuAlaSerSerLysSerSerThrAlaLeuSerPheThrLysProTh 113
QY 236 TGCC-----TGCGATGCCACTATGCTGGTTCGTACAGGTGA 274
Db 113 rAlaSerPheGlnAlaThrLeuHisCysTryGlyAlaTyLeuThrVaIGlyThrSerPr 133
QY 275 AACCTTGATGCTGAATTTGGCAGATCGAAAA-----AACATTATGGA 319
Db 133 oSerCysValValAsnPheAlaGlnLeuTrpLysProPheValThrArgAlaTySerGl 153
QY 320 AAAGATGAACGG 332
Db 153 uLysAspThrArg 157

RESULT 4
D71365
probable 64.7K tpr protein - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 11-Jan-2000
C:Accession: D71365; H71362
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: D71365
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-598 <COL>
A:Cross-references: GB:AE001196; GB:AE000520; NID:g3322372; PIDN:AAC65107.1; PID:g332
A:Experimental source: strain Nichols
A:Accession: H71362
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-598 <CO2>
A:Cross-references: GB:AE001198; GB:AE000520; NID:g3322390; PID:g3322392
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: tprC(TP0117); tprD(TP0131)
A:Superfamily: Treponema pallidum tprL protein

Alignment Scores:

Pred. No.: 4.07 Length: 598
 Score: 70.50 Matches: 37
 Percent Similarity: 35.86% Conservative: 15
 Best Local Similarity: 25.52% Mismatches: 35
 Query Match: 11.65% Indels: 38
 DB: 2 Gaps: 6

US-09-856-221-4 (1-332) x D71365 (1-598)

QY 53 TCAGGCAGCTGGCGCTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATAACTGGCT 112
 Db 19 SerGlyTyrAlaGlyValLeuThrProGlnVal-SerGlyThrAlaGlnLeuGlnTrp-- 37
 QY 113 ACGTGGCGGATGGCGCTATTATTATCAG----- 143
 Db 38 -----GlyIleAlaPheGlnLysAsnProArgThrGlyProGlyLysHi 52
 QY 144 -----TTTATGACTTGGCTGTTCCCTGTGTTGATGCTGAACA 184
 Db 52 stHrHisGlyPheArgThrAsnSerLeuThrIleSerLeuProLeuValSerIyshi 72
 QY 185 AACTTAC-----CAGTATGAATTAACGA 208
 Db 72 stHrHisThrArgArgGlyGluAlaArgSerGlyValTrpAlaGlnLeuGlnLysAs 92
 QY 209 TAAAGCTGTACGC-----TTCATTAAGCCGG 235
 Db 92 pLeuAlaValGluLeuAlaSerSerLysSerSerThrAlaLeuSerPheThrLysProTh 112
 QY 236 TGCC-----TGGCATGGCACTTATGCTGGTTGTGTAGCAGGTGA 274
 Db 112 rAlaSerPheGlnAlaThrLeuHisCysTyrGlyAlaTyrLeuThrValGlyThrSerPr 132
 QY 275 AACCTTGATGCTGAATTTGGCAGATGGAATA-----AACTATTGGA 319
 Db 132 oSerCysValValAsnPheAlaGlnLeuTrpLysProPheValThrArgAlaTyrSerGl 152
 QY 320 AAAGATGAACGG 332
 Db 152 ulysAspThrArg 156

RESULT 5
 F82207
 ATP-dependent helicase HrpA VC1382 [imported] - Vibrio cholerae (strain N16961 serogroup O1, R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.)
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: F82207
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B. Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: F82207
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1309 <HEI>
 A:Cross-references: GB:AE004217; GB:AE003852; NID:g9655866; PIDN:AAF94540.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC1382
 A:Map position: 1

Alignment Scores:
 Pred. No.: 4.31 Length: 1309
 Score: 70.50 Matches: 23
 Percent Similarity: 43.43% Conservative: 20
 Best Local Similarity: 23.23% Mismatches: 31
 Query Match: 11.65% Indels: 25
 DB: 2 Gaps: 4

US-09-856-221-4 (1-332) x F82207 (1-1309)

QY 18 CAAAAAACTATCTGGAACCCACAGCACAATACTCAGCAGCAGTGCCTTCCTACAA 77
 Db 579 GlnGlnLysAlaLeuThrSerAsnGlnPheArgGlnCysLysLeuAspTyrLeu--- 597
 QY 78 AGCAAAATTCAGCAATACAGCGTTGTATAACTGGCTAGCTGGCGCA---TTGCGCGCTATT 134
 Db 598 -----AsnTyrLeuArgValArgGluTrpGlnAspVal 608
 QY 135 TATTATCAGTTTATCAGTTCGCTGTTTCCCTGTGTTGATGCTCAACAACTTACCAG 194
 Db 609 TyrThrGlnLeuHis-----GlnSerThrArgGluMetGly 620
 QY 195 TATGAATTGAACATAAAGCTGTACGCTTCATTAAAGCCCGTGCCTGGCAGCTTAT 254
 Db 621 PheLysLeuAsnAsp-----GluProGlySerTyrHisAlaValHis 634
 QY 255 GCTGGTTGTAGCAGTGCAACCTTGATGCTGAATTTGGCAGCAGATGGAAAAAAC 311
 Db 635 SerAlaIleLeuValGlyLeuLeuSerHisIleGlyMetLysAspGlnGluLysAsn 653

RESULT 6
 A21185
 hypothetical protein - Trypanosoma cruzi (fragment)
 C:Species: Trypanosoma cruzi
 C:Date: 07-Sep-1990 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
 C:Accession: A21185
 R:Gonzalez, A.; Prediger, E.; Huecas, M.E.; Nogueira, N.; Lizardi, P.M. Proc. Natl. Acad. Sci. U.S.A. 81, 3356-3360, 1984
 A:Title: Minichromosomal repetitive DNA in Trypanosoma cruzi: its use in a high-sensi
 A:Reference number: A21185; MUID:84221958; PMID:6427769
 A:Accession: A21185
 A:Molecule type: DNA
 A:Residues: 1-194 <CON>
 A>Note: the authors translated the codon GAG for residue 15 as Gln

Alignment Scores:
 Pred. No.: 4.33 Length: 194
 Score: 70.00 Matches: 19
 Percent Similarity: 50.94% Conservative: 8
 Best Local Similarity: 35.85% Mismatches: 24
 Query Match: 11.57% Indels: 2
 DB: 2 Gaps: 2

US-09-856-221-4 (1-332) x A21185 (1-194)

QY 16 TGAATAAACTATCTGGAACCCACAGCACAATACTCAGCAGCAGTGCCTTCCTAC 75
 Db 45 CysValPheThrHisTrpThrProAsnAsnProGluLeuSerAlaAlaTrpArgAspPhe 64
 QY 76 AAAGCAAAATTCAGCAATACAGCGTTGTATAACTGGC---TACGTGGCGCATTTGGCGGCTA 132
 Db 65 ---AlaSerSerCysProGlnGlyCysCysThrArgLeuIleValPheGluTrpLeuLeu 83
 QY 133 TTTATTATCAGTTTATGACTTGGCTGTTTCCCTGTGTT 171
 Db 84 HisHisThrLeuTrpSerLysPheLeuPheProIleVal 96

RESULT 7
 E81828
 Probable periplasmic type I secretion system protein NMA1994 [imported] - Neisseria m
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: E81828
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
 A:Reference number: A81775; MUID:20222556; PMID:10761919
 A:Accession: E81828
 A:Status: preliminary
 A:Molecule type: DNA

A;Residues: 1-467 <PAR>
A;Cross-references: GB:AL162757; GB:AL157959; NID:q7380371; PIDN:CAB85213.1; PID:g738062
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NNA1992, NNA1994

Alignment Scores:
Pred. No.: 5.33 Length: 467
Score: 69.50 Matches: 30
Percent Similarity: 42.99% Conservative: 16
Best Local Similarity: 28.04% Mismatches: 38
Query Match: 11.49% Indels: 23
DB: 2 Gaps: 5

US-09-856-221-4 (1-332) x B81828 (1-467)

QY 33 GAACCCACAGCGCAACACTCAGGCAGCTGGCGCTTCCTACAAAGCAAAATTCAGCAAT 92
Db 176 AspThrValAlaAlaHisAlaAlaGluLysGluAlaTyraAlaGlnValArgGlnAla 195
QY 93 ACAGCGCTGTATAACTGCTACGTGGCGGATTTGGCGGCTATTATTATCAG----- 143
Db 196 GlnAlaLeuPheAsn-----LysGlyAlaAlaThrAlaLeuAspIleHisGluAlaLys 213
QY 144 -----TTTATGAC-----TTGGCTGTTTCCCTGCTGTTGCTGCTGAA---CAAACTTAC 191
Db 214 AlaGlyTyraAspAlaAlaLeuAlaGlnGluAlaValAlaLeuAlaGluLysGlnThrTyr 233
QY 192 CAGTATGAATTAACGATTAAGCTGTACGCTTCATTAAGCCCGGCTGCTGCGATGCACACT 251
Db 234 GluAsnGlnLeuAsnAsp----- 239
QY 252 TATGCTGGTGTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGACATGGAAAAAAC 311
Db 240 TyrThrGlyLeuAspSerLysGlnIleGluAlaIleAspThrAlaAsnLeuLeuAlaArg 259
QY 312 TATTGGAAAAAGATGAACGG 332
Db 260 TyrLeuProLysLeuGluArg 266

RESULT 8
NDECKR
Type I site-specific deoxyribonuclease (EC 3.1.21.3) EcoK chain R - Escherichia coli (st
A;Alternate names: type I restriction enzyme EcoK R chain; type I restriction-modificati
C;Species: Escherichia coli
C;Date: 31-Mar-1990 #sequence_revision 26-Sep-1997 #text_change 01-Mar-2002
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56576
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1188 <BURL>
A;Cross-references: EMBL:U14003; NID:q1263172; PIDN:AAA97247.1; PID:g537192
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H65249
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1188 <BLAT>
A;Cross-references: GB:AE000505; GB:U00096; NID:q2367375; PIDN:AAC77306.1; PID:g1790809;
A;Experimental source: strain K-12, substrain MG1655
R;Waite-Rees, P.A.; Keating, C.J.; Moran, L.S.; Slatko, B.E.; Hornstra, L.J.; Benner, J.
J. Bacteriol. 173, 5207-5219, 1991
A;Title: Characterization and expression of the Escherichia coli Mrr restriction system.
A;Reference number: A40368; MUID:91317743; PMID:1650347
A;Accession: S18776

A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-27 <NAI>
A;Cross-references: EMBL:X54198
R;Loenen, W.A.M.; Daniel, A.S.; Braymer, H.D.; Murray, N.E.
J. Mol. Biol. 198, 159-170, 1987
A;Title: Organization and sequence of the hsd genes of Escherichia coli K-12.
A;Reference number: A30375; MUID:88118919; PMID:3323532
A;Accession: A30375
A;Molecule type: DNA
A;Residues: 1-628, ECCKNRRHRHPGATYCAFRAGLPLYPVTRRRSDRPGAYSDHHPQRAGGSLQRAGRAH
A;Cross-references: GB:X06545; NID:g41751; PIDN:CAA29791.1; PID:g41752
C;Comment: This is one of three components (S, R, and M chains) of type I site-specif
ctors; it is the site-specificity determinant for the catalytic activities of the enz
C;Genetics:
A;Gene: hsdR

A;Map position: 99 min
C;Superfamily: type I site-specific deoxyribonuclease EcoK chain R
C;Keywords: ATP; DNA binding; hydrolase; nucleotide binding; P-loop; restriction modifi
F:489-496/Region: nucleotide-binding motif A (P-loop)
F:588-593/Region: nucleotide-binding motif B
F:592-595/Region: DEAH motif

Alignment Scores:
Pred. No.: 5.7 Length: 1188
Score: 69.50 Matches: 30
Percent Similarity: 37.82% Conservative: 15
Best Local Similarity: 25.21% Mismatches: 33
Query Match: 11.49% Indels: 41
DB: 1 Gaps: 5

US-09-856-221-4 (1-332) x NDECKR (1-1188)

QY 21 AAAAATATCTGAAACCCAA-----CAGGCACAAACTCAGCAGCAGCTGCCTTC 71
Db 169 LysGlnGlnGlnGlnGlnValArgGlnLysAlaGlnThrGlnAlaGluValGluAla 188
QY 72 CTACAAAGCAAAATTCAGCAATACAGCTGTGTATAACGGCTA----- 113
Db 189 GlnGlnGlnLysLeu-----ValAlaLeuAsnGlyTyrIleAlaIleLeuGluGlyLys 206
QY 114 -----CGTGGCGATTGGCGCTATTATTATTATCAGTTTAT 149
Db 207 GlnGlnGlnThrGluAlaGlnThrGlnAlaArgLeuAlaAlaLeuGluAlaGlnLeuAla 226
QY 150 GACTTGCGCTGTTCCCTGCTGTTGATGCTGAACAA-----ACTTACCAGTATGAA 200
Db 227 GluLysAsnAlaGluLeuAlaLysGlnThrGlnGlnGlnGluArgLysAlaTyrHisLysGlu 246
QY 201 TTGAACGATAAAGCTGTACGCTTCATTAAAGCCCGTGCCTGGCATGCCTTATGCTGCT 260
Db 247 IleThrAspGlnAlaIleLys----- 253
QY 261 TTGTTAGCAGCTGAACCTTGATGCTGAATTTGGCAGACAGATGGAAAAAACTATTTG 317
Db 254 -----ArgThrIleuAsnLeuSerGluGluGluSerArgPheLeu 266

RESULT 9
A43336
microtubule-vesicle linker CLIP-170 - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C;Accession: A43336
R;Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A;Title: CLIP-170 links endocytic vesicles to microtubules.
A;Reference number: A43336; MUID:92405160; PMID:1356075
A;Accession: A43336
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1392 <PIE>
A;Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

Db 196 aserProHisPheAsnProGlnSerSerGluLysSerPheAsnGlnAspTyrAsnThrVa 216
 QY 235 -----GTGCTGGCAGTCACTTATGCTGTTTGTAGCAGGTGAACCTTGATGCTG 287
 Db 216 lAspGluLeuPro----- 220
 QY 288 AATTGGCAGATGGAAAAAACTATTGGAAAAAG 324
 Db 221 ----trpTyrLysTrpLysGlyHisValTrpAsnLys 231
 RESULT 12
 S75950
 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S75950
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 s.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S75950
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-258 <KAN>
 A:Cross-references: EMBL:D64006; GB:AB001339; NID:gl001291; PID:BAA10797.1; PID:d101144
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Alignment Scores:
 Pred. No.: 6.79 Length: 258
 Score: 68.50 Matches: 23
 Percent Similarity: 37.50% Conservative: 13
 Best Local Similarity: 23.96% Mismatches: 29
 Query Match: 11.32% Indels: 31
 DB: 2 Gaps: 5

US-09-856-221-4 (1-332) x S75950 (1-258)
 QY 57 GCACGCTGGCTTCTTCAAGCAAAATTCAGCAATACAGCGTGTATACATGGCTACGT 116
 Db 186 AlaGlnAlaThrHisLeuSerAlaSerPheAlaGlnThrLysAlaPheAsnTrpLeuGln 205
 QY 117 GGGCGATGGCGCTATTATATATACATTTATGACTTGGCTGTTCCCTGCTGTTGATG 176
 Db 206 AsnAsnAlaAla----- 209
 QY 177 GCTGAACAACTTACCAGTATGAATTG-----AAGATAAAGCTGTACGC 221
 Db 210 -----LysTyrSerPheGluLeuSerPheProAspAsnProGlnGlyIleAla 226
 QY 222 TTCATTAGCCGGCTGCTGCATGCGACATGCTGTTGTTAGCAGGTGAACCTTG 281
 Db 227 Tyr---GluPro-----TrpHisTrpArgTyr-----ValGlyAspArgGln 239
 QY 282 ATGCTGAATTGGCAGATGGAATAAACTATTGGAAAAAGATCAA 329
 Db 240 SerLeuGluLeuPheTyrLysAlaArgAsnLeuProGlnLysAsnGlu 255
 RESULT 13
 T33215
 hypothetical protein T07H8.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33215
 R:Gattung, S.; Maggi, L.
 submitted to the EMBL Data Library, May 1998
 A:Description: The sequence of *C. elegans* cosmid T07H8.
 A:Reference number: Z21303
 A:Accession: T33215
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA

A:Residues: 1-312 <GAT>
 A:Cross-references: EMBL:AF067945; PIDN:AAC17679.1; GSPDB:GN00023; CESP:T07H8.5
 A:Experimental source: strain Bristol N2; clone T07H8
 C:Genetics:
 ! A:Gene: CESP:T07H8.5
 A:Map position: 5
 A:Introns: 48/3; 126/3; 236/3; 282/3

Alignment Scores:
 Pred. No.: 6.88 Length: 312
 Score: 68.50 Matches: 35
 Percent Similarity: 41.67% Conservative: 15
 Best Local Similarity: 29.17% Mismatches: 30
 Query Match: 11.32% Indels: 40
 DB: 2 Gaps: 5

US-09-856-221-4 (1-332) x T33215 (1-312)
 QY 9 GCAGTATTGCAAAAAAATCTATCTG---GAAACCCACAGCAGCAGCACTCAGCAGCTG 65
 Db 155 AlaTyrLeuGluTyrAsnProLeuSerAspThrTyrValAlaArgThrGlnAlaAspLeu 174
 QY 66 GCCTTCCTACAA----- 77
 Db 175 ProPheLeuTyrSerPheIleLeuValTrpMetValValThrValLeuLeuSerIleIle 194
 QY 78 -----AGCAATTCAGCAATACAGCGTGTATACCTGG 110
 Db 195 AlaAsnIleIleCysTrpPheLysIleSerLysTyrSerLysAlaAlaArgGln----- 212
 QY 111 CTACGTGGCGATGGCGCTATTATTATATCAGTTTATGACTTGGCTGCTTCCCTGCT-- 168
 Db 213 ---Gln-SerAspTyrArgLeuPheLeuValSerPheValThrPheValIleAsnCysGln 231
 QY 169 -----GTTTGATGGCTGCAACAACTTACAGTATGAATTGAACGATAAAGCTGTA 218
 Db 231 yValPheSerIleAlaMetLeuAsnLysIleSerAlaAspIleAspProSerLysLeuLe 251
 QY 219 CGCTTCA-----TTAAGCCGCTGCTGGCATGGCATGCTTATGCTG 258
 Db 251 uLeuSerSerArgIleAlaGlnLeuLeuSerProPheAlaAsnAspLeuLeuSerLeu 270
 RESULT 14
 AH0159
 probable lysine-specific permease *lysp* [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AH0159
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
 deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AH0159
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-503 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90139.1; PID:gl5979359; GSPDB:GN00175
 C:Genetics:
 A:Gene: *lysp*
 C:Superfamily: arginine permease

Alignment Scores:
 Pred. No.: 8.22 Length: 503
 Score: 68.00 Matches: 35
 Percent Similarity: 35.07% Conservative: 12
 Best Local Similarity: 26.12% Mismatches: 29
 Query Match: 11.24% Indels: 58
 DB: 2 Gaps: 8

US-09-856-221-4 (1-332) x AH0159 (1-503)

